AD-758 875

CAPABILITIES OF GROUND OBSERVERS TO LOCATE, RECOGNIZE, AND ESTIMATE DISTANCE OF LOW-FLYING AIRCRAFT

Robert D. Baldwin

Human Resources Research Organization

Prepared for:

Office, Chief of Research and Development

March 1973

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

Technical Report 73-8

HumRRO-TR-73-8

HumRRO

AD 758875

Capabilities of Ground Observers to Locate, Recognize, and Estimate Distance of Low-Flying Aircraft

Robert D. Baldwin

HUMAN RESOURCES RESEARCH ORGANIZATION
300 North Washington Street • Alexandria, Virginia 22314

March 1973

Prepared for

Office of the Chief of Research and Development
Department of the Army
Washington, D.C. 20310

59

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Human Resources Research Organization (HumRRO) 300 North Washington Street Alexandria, Virginia 22314	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

CAPABILITIES OF GROUND OBSERVERS TO LOCATE, RECOGNIZE, AND ESTIMATE DISTANCE OF LOW-FLYING AIRCRAFT

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5. AUTHOR(S) (First name, middle initial, last name)

Robert D. Baldwin

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
March 1973	56	19

8a. CONTRACT OR GRANT NO. DAHC 19-73-C-0004 b. PROJECT NO. 2Q062107A745 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) HumRRO TR-73-8 9b. OTHER REPORT NO.(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
HumRRO Division No. 5, Fort Bliss, Tex.; Basic Research 16, Improving Ability to See Military Targets	Office, Chief of Research and Development Department of the Army Washington, D.C. 20310

13. ABSTRACT

A considerable amount of research has been conducted during the past 10 years concerning the abilities of ground observers to detect, recognize, and estimate the range of aircraft. This report integrates and evaluates the results of 20 technical reports concerning these abilities. The effects on visual detection and recognition of visual aids, search sectors, target altitudes, and search methods are discussed. The techniques used for training in aircraft recognition are reviewed for historical origins and research validity. The accuracy of ground-to-air range estimation is described for unaided and stadimetric (size-distance) methods. Techniques of training range estimation in miniaturized situations are described. The influence of environmental factors on estimation accuracy is examined.

DD FORM 1 NOV 65 1473

Unclassified

Security Classification

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Aircraft recognition						
Aircraft detection						
Distance estimation						
Ground observers						
Perceptual training						
Range estimation						
Stadimetric Aids						
Visual detection						
Visual perception						

Unclassified

Security Classification

HumRRO
Technical
Report
73-8

Capabilities of Ground Observers to Locate, Recognize, and Estimate Distance of Low-Flying Aircraft

Robert D. Baldwin

HumRRO Division No. 5
Fort Bliss, Texas

HUMAN RESOURCES RESEARCH ORGANIZATION
300 North Washington Street • Alexandria, Virginia 22314

Basic Research Study 16

March 1973

Approved for
public release;
distribution
unlimited.

Prepared for

Office of the Chief of Research and Development
Department of the Army
Washington, D.C. 20310

The Human Resources Research Organization (HumRRO) is a nonprofit corporation established in 1969 to conduct research in the field of training and education. It is a continuation of The George Washington University Human Resources Research Office. HumRRO's general purpose is to improve human performance, particularly in organizational settings, through behavioral and social science research, development, and consultation. HumRRO's mission in work performed under contract with the Department of the Army is to conduct research in the fields of training, motivation, and leadership.

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

Published
March 1973
by
HUMAN RESOURCES RESEARCH ORGANIZATION
300 North Washington Street
Alexandria, Virginia 22314

Distributed under the authority of the
Chief of Research and Development
Department of the Army
Washington, D.C. 20310

FOREWORD

Since 1963, the Human Resources Research Organization has conducted extensive research on the perceptual skills required of operators of man-ascendant air defense weapons. Such weapon systems are visually sighted and frequently have no auxiliary electronic systems for use in identifying or estimating the range of low flying aircraft.

Following on preliminary Exploratory Studies, a majority of the HumRRO research on the visual detection, recognition, and ranging of low-flying aircraft was accomplished under Work Units STAR and SKYFIRE. A number of HumRRO Technical Reports documented various phases of these extensive programs of applied research on perceptual performance and training.

The present report was prepared as a source document that summarizes most of the HumRRO research, as well as contemporary applied research conducted by other human factors groups in these areas. It encompasses all the known contemporary unclassified research on the subject.

The research was performed at HumRRO Division No. 5, Fort Bliss, Texas. The author, Dr. Robert D. Baldwin, was Director of Division No. 5 during the 1962-1970 period when most of the research reported here was accomplished. Dr. Albert L. Kubala is the present Director of the Division. Dr. Baldwin prepared the present document, with critical reviews contributed by Dr. Kubala, Dr. Elmo E. Miller, and Dr. Paul G. Whitmore, all of Division No. 5.

Military support throughout the period of these research projects was provided by the U.S. Army Air Defense Human Research Unit. LTC Frank D. Lawler is the present Military Chief.

HumRRO research for the Department of the Army is conducted under Contract DAHC 19-73-C-0004. Army Training Research is conducted under Army Project 2Q062107A745.

Meredith P. Crawford
President
Human Resources Research Organization

SUMMARY

BACKGROUND

During the past 10 years, there has been a substantial investment in research concerning the abilities of ground observers to detect, identify, and estimate the distance of low-flying aircraft. In that period, extensive programs of research on perceptual abilities have been conducted at HumRRO Division No. 5 for the U.S. Army, and a number of technical reports documenting the various phases of these programs have been published. The present report was prepared as a source document, summarizing most of this HumRRO work, as well as the research conducted in these areas by other human factors research groups. These experimental and analytical studies originally were described in separate, isolated reports. The purpose of this summary presentation is to integrate and evaluate the available unclassified information concerning (a) visual detection performance, (b) aircraft recognition ability and training methods, and (c) visual ranging ability against low-flying aircraft.

VISUAL DETECTION

Five major field experiments were conducted during 1960-1965 in the Southwest United States and in Germany to evaluate man's ability to visually detect low-flying jet aircraft. These experiments used varied periods of early warning, different sizes of search sector, and different conditions for aided and unaided viewing.

Collectively, these experiments indicated that there was a strong deterministic relationship between the range at which an aircraft was detected and the accuracy of the early warning data provided. Detection ranges of less than 2,000 meters occurred when search sectors of 180° to 360° were used and no information was provided concerning expected time of appearance. At the other extreme, detection ranges exceeding 12,000 meters occurred when 5° sectors were used and accurate information concerning "attack" time was available.

The use of hand-held binoculars during surveillance tasks did not facilitate detection. In fact, when nearby terrain features blocked the view of the distant horizon, detection with visual aids occurred later than when unaided search was used.

Aircraft altitude and the observer's location also influenced detectability. Aircraft at 500 feet tended to be seen sooner than those at 1,500 feet and observers who were offset from the flight path tended to make detections sooner.

The limited research on search techniques has yielded equivocal results. Attempts to teach observers specific structured or systematic methods of sky search aided some persons, but hampered the effectiveness of others. There was a suggestion from the data that instruction on systematic search methods did increase the effectiveness of persons with average or below average visual efficiency. However, such training may have interfered with established searching patterns developed by people with highly efficient visual systems.

VISUAL RECOGNITION ABILITY

Research concerning aircraft recognition ability consists of two full-scale studies using a small number of actual aircraft and two reduced-scale or miniaturized field studies

using a larger variety of model aircraft. These studies were concerned with evaluating man's accuracy in recognizing aircraft, and with determining the distance to the aircraft at the time it is recognized.

The full-scale tests were accomplished in conjunction with visual detection tests. As a result, the distances at which the aircraft were recognized were affected by the conditions established for visual detection as well as the recognition skills of the observers. One full-scale test, which used large search sectors and no early warning, obtained average recognition ranges of about 900 meters. On the other hand, when small search sectors and fairly accurate early warning were employed in another test, the average recognition range was 4,000 meters.

A miniaturized test, which used about a dozen different aircraft models, obtained even greater recognition ranges—as great as 10,000 meters for some multiengined aircraft. Although "professional" observers were used, the results of that test, coupled with the results of the full-scale tests, suggest that recognition range varies with recognition accuracy. That is, the more highly motivated and trained observers tend to recognize aircraft sooner as well as with greater accuracy. This inference from the collective results of all the tests seems to run counter to a common-sense expectation that accuracy should be inversely related to aircraft distance.

Another miniaturized field test evaluated the recognition performance of crew chiefs working with and without the assistance of forward observers. This test also provided interesting but conflicting results. The forward observer's preliminary judgments concerning an aircraft's identity seemed to facilitate the recognition decisions of some crew chiefs, but interfered with the decisions of others. In many instances, crew chiefs made their decisions *before* receiving the supposedly advanced input from the forward observer.

Optical aids have been found to increase the distance at which an aircraft is recognized. Even though optics may interfere with initial detection, they do increase the visibility of the distinguishing features by which aircraft are discriminated.

AIRCRAFT RECOGNITION TRAINING

The principal aircraft recognition training method available in the early 1960s was basically unchanged from that used by U.S. armed forces in World War II. The method consisted of a dubious mixture of techniques based upon diametrically opposed concepts concerning shape for form discrimination learning. One concept, proposed in the 1940s by Samuel Renshaw, emphasized learning to recognize whole shapes or images that were presented to students for very brief time intervals by means of tachistoscopes. A second, opposed, concept emphasized the learning of the component parts of aircraft. Called "WEFT," it employed the technical terms used by the aviation industry for describing the Wings, Engines, Fuselage, and Tail assemblies.

Although research results refuting the claimed advantages of the Renshaw "whole image" approach were available during World War II, they were largely buried in retired files and were only recently brought to light again. However, in the mid-1960s, HumRRO psychologists began evaluations of alternative methods for teaching aircraft recognition. These evaluations indicated that the mixture of concepts that had evolved from experience in World War II did not produce the level of recognition proficiency desired by the armed forces.

As a result of extensive applied research, a set of techniques and training aids were identified that would produce the desired proficiency levels—the Ground Observer Aircraft Recognition (GOAR) method. The GOAR techniques emphasize the initial learning of aircraft features that distinguish one shape from another, followed by discrimination training in which pairs of similar aircraft are viewed simultaneously, and culminating in recognition practice with single images presented successively using a stimulus-response-feedback paradigm. Throughout this learning cycle, esoteric technical descriptors and extremely short viewing intervals are avoided.

Additional research was performed to evaluate training methods for self-study use that used printed images rather than optical projection of aircraft forms. This research indicated that the GOAR techniques originally developed for classroom use were also optimal for self-study applications.

RANGE ESTIMATION

Much of the past research on distance estimation has been concerned with judging the range to ground-based objects. Very little was known about observers' accuracy in judging the distance to moving aerial objects. An extensive series of studies was begun by HumRRO in 1965 on measuring estimation accuracy, evaluating training methods, using simple stadimetric aids for ranging, and identifying factors that affect accuracy in judging the sizes and distances of objects.

Training techniques that used cueing and knowledge of results were evaluated, but they were found to produce less accuracy than was obtained by using stadimetric aids in an "open fire" event that was judged by comparing the apparent size of the target with the size of a reference or stadia rod.

Field tests of ranging errors without stadia references indicated that judgmental errors varied with the distance to be estimated.

For distances beyond 3,000 meters, observers judged the aircraft to be more distant than was actually the case (an overestimation); however, for distances less than 1,400 to 1,500 meters the aircraft range tended to be underestimated (i.e., aircraft was erroneously judged as nearer).

Factors such as observer offset, aircraft altitude, and illumination were evaluated. In one study, error magnitude decreased as offset increased. Although illumination did not affect accuracy, target altitude did. Error magnitude both with and without stadimetric aids was not equal for incoming and outgoing directions.

Considerable research was done on the use of miniaturized training facilities in which 1:50-scale model aircraft were moved toward and away from the observers. Most of this research concerned the use of stadimetric ranging aids for estimating open and cease-fire events. These experiments indicated that reduced-scale training was effective when measured by stadimetric accuracy in a full-scale environment involving actual aircraft.

Retention of stadimetric ranging skill 30 days following reduced-scale training also was of acceptable accuracy for judging the "open fire" events on inbound aircraft, but errors of 200 meters in judging a 1,500-meter "cease fire" event occurred. This tendency for errors to be greater for outbound than inbound flights was found in most of the field tests. Research is currently under way in HumRRO Basic Research Program 16 to identify the factors that cause errors in stadimetric ranging.

CONTENTS

Figures *(Cont.)*

Capabilities of Ground Observers to Locate, Recognize, and Estimate Distance of Low-Flying Aircraft

INTRODUCTION

In the past decade, there has been a substantial investment in research on the perceptual abilities of ground observers to detect, identify (or recognize), estimate the range of, and engage low-flying aircraft.

In order to escape detection and engagement by radar-controlled air defense systems, aerial attack tactics have emphasized the use of low altitude penetration and bombardment procedures. To counter such attack procedures, the U.S. Army established requirements for the development and deployment of low altitude air defense systems. Because contemporary radar systems are unable to discriminate aircraft from adjacent terrain, the new low altitude air defense systems were dependent upon the human operator for the localization and identification of attacking aircraft. Many of these systems also depended upon man's judgment as to when the aircraft was within the engagement envelope of the defending weapon.

With the advent of such low altitude weapons, military planners and training agencies needed information concerning the limits of man's ability to accomplish the required perceptual tasks. Operations research organizations, both civilian and military, needed a data bank from which to draw information concerning human capabilities in this area. Military trainers needed to know the limits of human perceptual ability, and to acquire information on techniques for maximizing those human abilities in the detection, recognition, and range estimation tasks.

For the most part, the experimental and analytical studies that resulted from these requirements have been reported in separate technical reports issued by various in-service and contract agencies. Each individual experiment or study, therefore, tends to stand in isolation. The purpose of the present report is to integrate and evaluate the information collected during the past 10 years concerning the perceptual skill and capabilities of ground-based observers.

Chapter 1

VISUAL DETECTION

BACKGROUND

The sequence of actions that culminates in a decision to engage an incoming aircraft begins with its detection and/or localization. The predominant emphasis in the perceptual skills experimentation concerning gound observers has been upon visual detection. This emphasis on detection, rather than on recognition, ranging, or other skills involved in aircraft engagement, appears to have grown out of the predominant interests of military operations research groups in developing models to assess the effectiveness of alternative air defense weapons attempting to engage high performance aircraft.

The interest of war game and simulation specialists in visual detection no doubt resulted from their efforts to predict the technical capabilities and requirements of air defense weapons that would at least match the outer bounds of man's capabilities to engage aircraft. Since the maximum engagement range for an air defense weapon is determined by the maximum detection range of the human observer, there were more requirements for experimentation and research on detection than on the other perceptual skills. The other perceptual skills, such as aircraft recognition, friend-foe identification, and range estimation, undoubtedly occupied less attention on the part of military planners and analysts because of the assumption that aids to recognition (such as electronic interrogation equipments) and aids for determining open-fire range (such as shorter range radar equipments) would reduce somewhat the burden on the human observer for accomplishing these functions.

FIELD TESTS

THE GILA BEND TESTS

With the exceptions of one classified Department of Defense study and an experiment conducted in Germany, all visual detection experimentation has been performed in a desert environment. The earliest of the recent studies on detection was conducted by the U.S. Army Human Engineering Laboratories (HEL), as reported by Wokoun (1). The HEL study was concerned with the effectiveness of the ground observer in detecting and identifying jet aircraft, as a function of the size of the sector being searched and the aircraft altitudes.

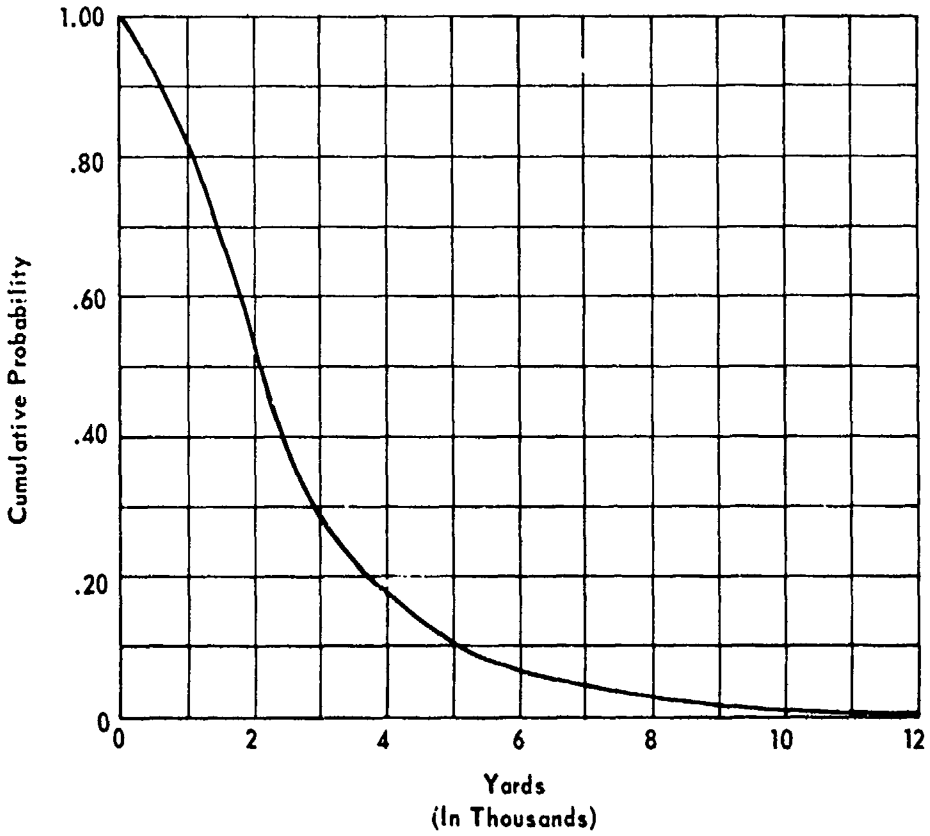
The HEL study was conducted at Gila Bend, Arizona over a five-day period in September 1959. It involved unaided detection of aircraft with no provision for early warning either temporally or spatially. At the time, metrorological visibility was greater than 15 miles and the aircraft were presented against a clear sky. Wokoun reported data for 15 subjects. All observers had visual acuity of 20/22 or better, and were between the ages of 19 and 25. Targets to be detected were T-33, F-86, and F-100 jet fighters. These aircraft flew along six courses that were randomly scheduled during the tests and at two altitudes, 500 and 1,500 feet, which were also randomized over trials. Aircraft flew at a

Preceding page blank

speed of approximately 400 knots. Prior to the test, observers were trained in the use of a vertical scan or saw-tooth searching technique employing the far horizon as a touch point to avoid empty field myopia. Only the detection data will be reported in this chapter; the aircraft identification results will be presented later.

Four search sectors were used: 45°, 90°, 180°, and 360°. When the results were summed over all four search sectors, 50% of the detections occurred when the aircraft was at least 2,050 yards distant. A considerable number of interactions or variability in the data occurred as a function of the width of the search sector and the target altitude. In general aircraft at 500 feet were detected earlier than at 1,500 feet. There was some suggestion that training or experience in the field situation tended to improve performance after one day of practice. In addition, as the crossover range increased, detection range tended to increase (although not invariably), apparently because of the greater visual angle presented by the aircraft as the aspect moved from head-on to tangential. Figure 1 presents a summary of all Wokoun's field experimentation, showing the cumulative percent of aircraft detected as a function of the distance of tne aircraft from the observer. The narrower search sectors, 45° and 90°, produced earlier detections than the larger search sectors, 180° and 360°. The combination of 45° search sector and aircraft altitudes of approximately 500 feet produced the earliest detections.

Cumulative Probability That An Incoming Aircraft Will Be Detected by Given Distances From Observer When All Search Sector Sizes and Both Altitudes Are Combined



NOTE: From Wokoun (1).

Figure 1

THE WHITE SANDS TEST

The next major field experiment was conducted in October 1961 at White Sands Missile Range (WSMR) north of El Paso, Texas (Zimmer and McGinnis, 2). Although preliminary instruction was given in vertical scan procedures, using the horizon as a reference boundary, the 24 observers were told to use any scanning technique that personally seemed to be effective. All observations were made unaided—that is, no optical aids were used. Three types of aircraft were used: propeller, helicopter, and jet (T-33 and F-100). The jet aircraft flew speeds varying between 200 and 400 knots.

Three alert conditions were established for the observations. In one condition, all observers were assigned a 30° search sector and were given 15 minutes warning concerning the approach of an aircraft. In the second condition, the search sector was 180°, and similar warning data were provided. In the third, the search sector was 180°, but no time of arrival or approach information was given.

Approximately 2,200 observations were obtained during the test period. Summing over all observation conditions, the mean detection occurred at 5,130 yards, with a standard deviation equal to 3,177 yards. Ninety percent of the detections occurred between 1,200 and 11,200 yards. Although aircraft altitude varied between less than 150 feet to 3,200 feet, no consistent results concerning effects upon detection range occurred. The data did suggest, however, that detection occurred sooner if the aircraft altitude was between 150 and 1,200 feet. Altitudes below and above those limits tended to produce later detections. The degree of alert established produced no consistent effects upon the detection ranges obtained, although the data suggest that increasing the amount of early warning tended to increase the detection range. These results, however, were not statistically significant.

The cumulative frequency of detection, as a function of aircraft distance, is shown in Figure 2 for the F-100 aircraft. Fifty percent of the detections occurred prior to or by the time the aircraft had approached within 3,500 meters.

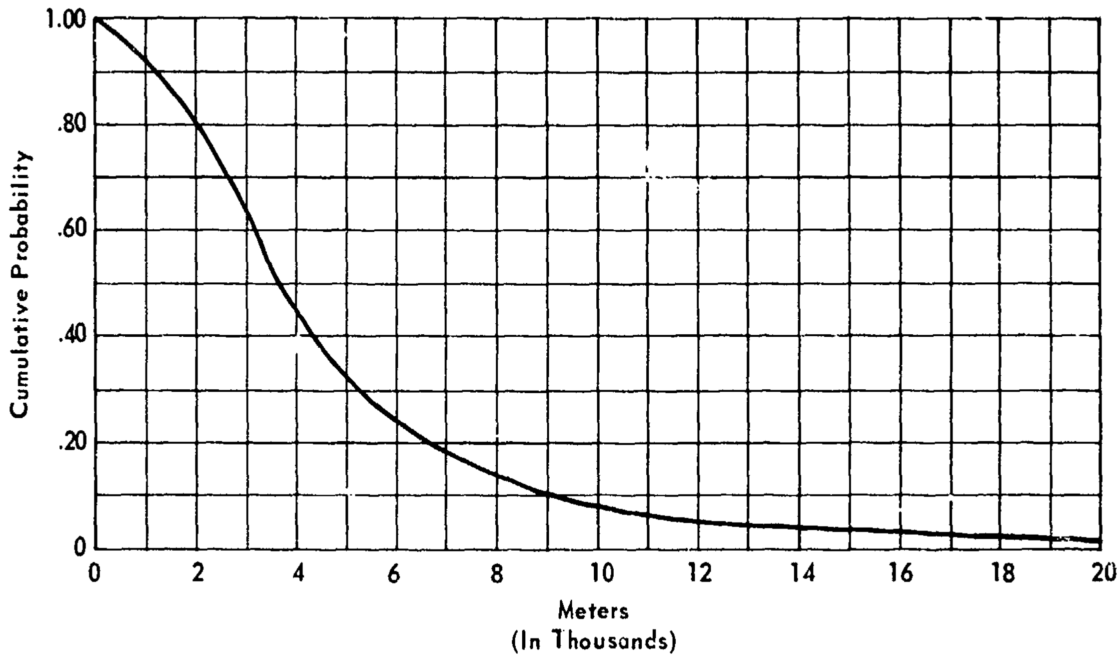
THE DONA ANA TEST

In April 1965, a field experiment was conducted by HumRRO (Wright, 3) at Dona Ana Range Camp, Fort Bliss, Texas, which employed the same general terrain as used for the White Sands test. The terrain allowed for low altitude aircraft approaches up to 20 miles in length which were unobstructed by terrain masking. The principal objective of the HumRRO field study was to determine man's unaided ability to visually detect and recognize low altitude aircraft under optimum field conditions with respect to early warning. Only the detection data will be discussed in this chapter. In this field experiment, the effectiveness of optical aids was also evaluated.

During the tests, visibility was greater than 90 miles for every test day, except one, when it was not less than approximately 50 miles. Twenty-seven enlisted men served as observers for the tests. All observers had visual acuity of 20/25 or better as measured by the Armed Forces Vision Tester. As part of preliminary target detection training, all observers were given training in search techniques, which consisted of practice in observing a horizontal search scan with frequent orientation to distant terrain on the horizon. The observers who were to employ binoculars were trained in their use for search and in techniques for holding binoculars steady.

Prior to the measured field test, the observers viewed 27 jet, 15 propeller, and 10 helicopter passes before formal data collection began. The targets consisted of F-4C, F-100, and T-33 jet aircraft, which flew at a speed of approximately 400 knots at 100-300 feet altitude. Propeller aircraft targets consisted of a 0-1A, U-6A, and U-1A,

Detection Range for F-100 Aircraft



NOTE: After Zimmer and McGinnis (2).

Figure 2

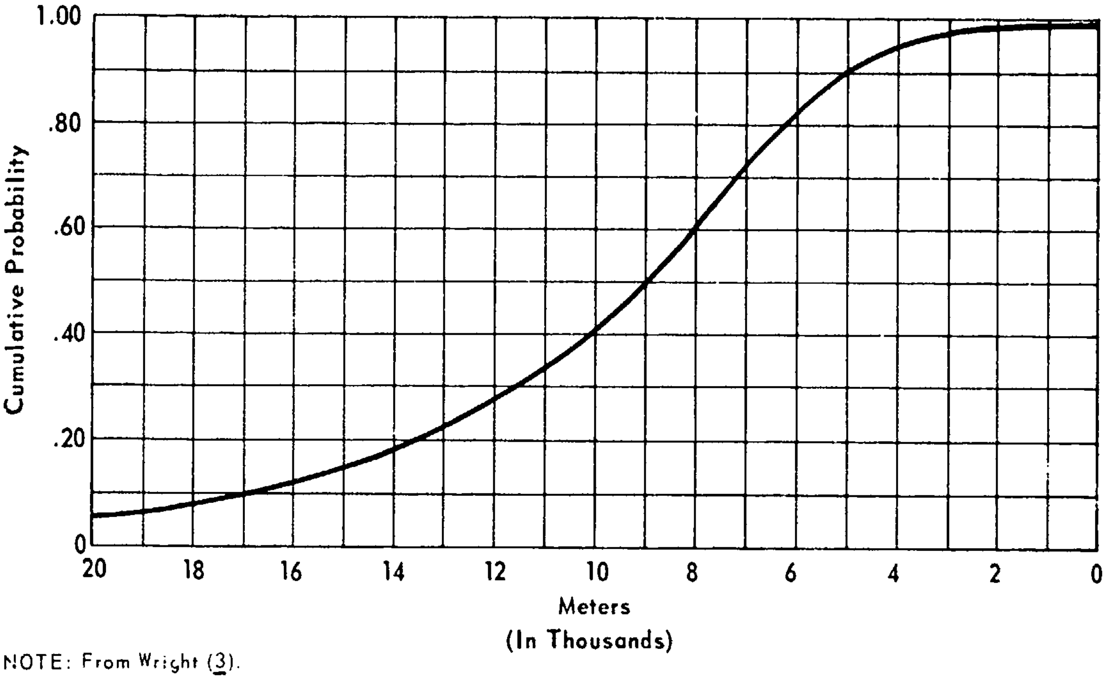
which flew at speeds of approximately 100 knots and at altitudes between 100 and 300 feet. Data were also obtained on one OH-23 helicopter, which flew at a speed of about 75 knots at 100 feet or less.

During this test, 18 of the observers had no visual aids for aircraft detection, while nine observers were issued 6X30 binoculars. Observers were stationed in three clusters, with one group at the test control center over which all flight paths converged, another approximately 1,000 meters southeast of the test center, and the third about 2,000 meters north of the test center. Ten different flight paths were flown by the various aircraft. At the start of each aircraft pass, monitors informed the observers that an aircraft was inbound at a given clock position from their location.

A number of controlled factors were evaluated in the analysis, such as the use of visual aids, the amount of offset, the aircraft type, and all interactions of these main factors. As in prior field experiments, a considerable number of interactions occurred among the variables. For those observers using unaided vision, the average distance of the aircraft at the time of detection was 10,700 meters. For those equipped with 6X30 binoculars, the mean detection range was approximately 11,900 meters. The use of optical aids tended to interact with the offset of the observers from the minimum crossing range of the aircraft. Unaided detection was earlier than aided detection for the head-on or incoming aircraft. In contrast, those observers equipped with the binoculars tended to detect aircraft sooner than the unaided observers when the aircraft was flying a tangential or offset course.

Figure 3 presents the cumulative probability curves for Wright's experiment, summed over all aircraft and all conditions that were used. It can be seen that visual detection occurred with a 50% probability by the time the aircraft had at least approached within 9,000 meters of the observer.

Empirical Probabilities of Detection



NOTE: From Wright (3).

Figure 3

The detection probabilities for each of the three jet aircraft are presented in Figure 4. As shown, the size of the aircraft appears to influence the likelihood that it will be detected. The F-4C, the largest of the three aircraft used, yielded a 50% probability of detection at a distance greater than 14,000 meters. In contrast, the T-33, which was the smallest aircraft used, had a comparable detection likelihood at only 7,000 meters.

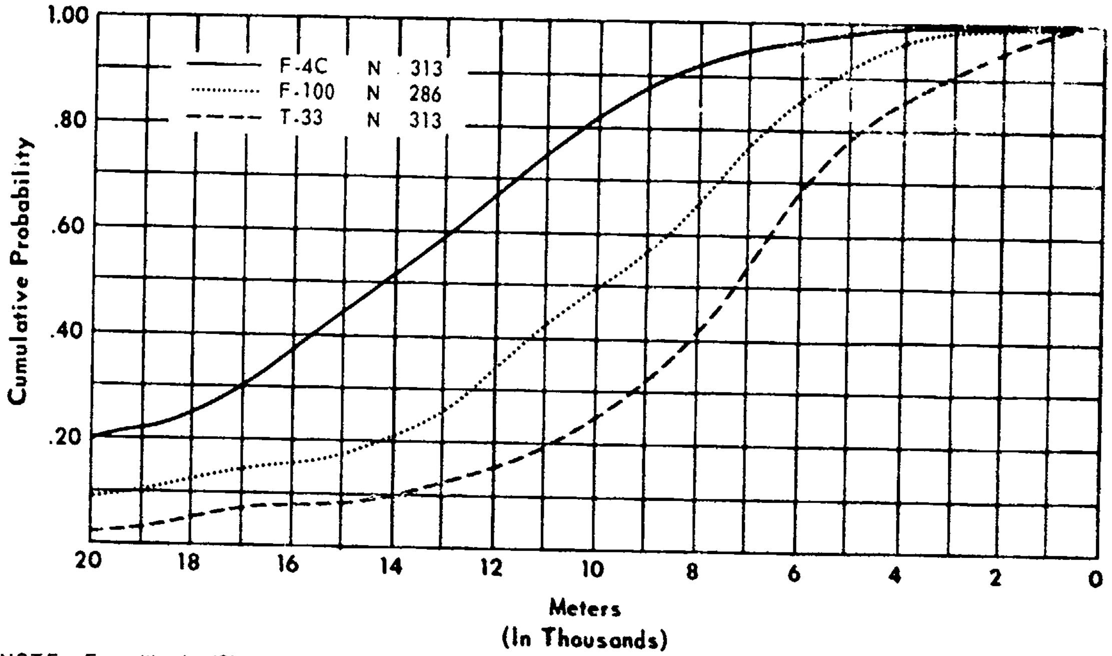
THE TONOPAH TEST

A second aircraft detection test was conducted by HumRRO during June and July 1965, near Tonopah, Nevada (1). A portion of the Tonopah tests was performed to obtain data concerning the effects on visual detection of (a) the amount of lateral offset of the observers from the flight path of the aircraft, (b) the type of visual aids used and their optical power, and (c) the amount of early warning given the observers. In part, these tests were motivated by the results obtained earlier in the year by Wright, who found detection ranges considerably greater than those reported in the earlier tests. The eight observers used for these tests were military and civilian research personnel who had been involved in the earlier detection experiment.

The aircraft flight path was over a wide, flat desert valley between two lines of barren mountains that began rising about 500 feet to the west and 1,500 feet to the east of the flight path's ground projection. The test site was adjacent to the flight path on relatively flat terrain. The observers were located at four observation posts (OPs) at distances of 200, 1,100, 2,600, and 3,300 meters perpendicular to the flight path.

The aircraft flew an essentially north-to-south or south-to-north pattern. To the south, the view of the aircraft was interrupted by the distant horizon approximately 15

Effect of Aircraft Range and Aircraft Type on Detection Probabilities



NOTE: From Wright (3).

Figure 4

miles away. As one moved from the near OP (200 meters) to the most distant OP (3,300 meters), the near terrain east of the flight path increasingly became the background for the flight path of the aircraft. The view of aircraft flights from the north was almost as good as that for those from the south, if the observer was located at 200 or 1,400 meters from the flight path.

At the time of target "unmask" (when it first became visible above the near horizon), the aircraft had a sky background when viewed from the 200-meter OP. For the other OPs, the aircraft had a terrain (distant mountain) background at the time of unmask. For those observers stationed at the 2,600- and 3,300-meter OPs, a ridge intervened between observer and aircraft on a north-to-south flight. As a result, the observers at these positions were not able to see the aircraft until it was much nearer to crossover (the intersection of the flight path and the observer's line of positions) than was the case at the nearer OPs.

Visibility during the test was never less than 40 to 50 miles. On four of the 20 test days, there were varying amounts of cloud cover. Sky brightness during the testing varied from approximately 1,500 foot-lamberts in the early morning to 3,400 foot-lamberts in the late morning.

One of the tests was a comparison of unaided observation vs. the use of 6X30 binoculars for aircraft detection. An F-4C aircraft was used for this test. The average range of the aircraft at the time of detection was 11,900 meters for unaided observation, and 12,200 meters when 6X30 binoculars were used. This difference was not statistically significant.

A third test evaluated aircraft detection range as a function of the amount of early warning provided the observers. Two levels of early warning were used: 1 minute and 5

minutes, prior to aircraft appearance on the horizon. This test used the F-105 aircraft, which flew courses involving both near and distant terrain masking. Under distant masking conditions, the mean detection ranges, when averaged over all OPs, were 12,150 meters for the 1-minute early warning condition, and 12,750 meters for the 5-minute early warning. Under near masking conditions, the detection ranges were 4,600 and 5,000 meters for the 1-minute and 5-minute warning, respectively. The effect of variation in amount of early warning was not statistically significant in either masking condition.

VISUAL AIDS FOR DETECTION

An earlier study was conducted in this area by the Human Engineering Laboratories, as reported by Kurke and McCain (5). This experiment compared binoculars, of from three to seven power, which were mounted on a pedestal and directed toward the line of appearance of the aircraft. Kurke and McCain found that the average detection ranges varied fairly regularly from 14,500 to 18,000 yards as optical power increased.

In contrast with the Kurke and McCain study, tests reported by Wright (3) and by Frederickson *et al.* (4) used hand-held binoculars to search the horizon. As discussed earlier, Wright found that the relative effectiveness of the unaided vs. aided observation varied with the offset of the observers from the flight path. For the "most threatening" targets (i.e., those with a head-on approach), unaided vision surpassed the aided observation. For the more oblique observation posts, binoculars tended to facilitate detection. Wright attributed this finding to the smaller field of view of the aided observation coupled with the relatively small subtended angle of the aircraft in the head-on position.

However, the Tonopah tests reported by Frederickson *et al.* (4) did not find any advantage for aided over unaided observation. In addition, there was no advantage in using higher powered (7X50) instead of slightly lower powered (6X30) binoculars.

The results of these tests tend to be confirmed by field experiments conducted in Germany (Doetsch and Hoffmann, 6), which compared the detectability of low-flying aircraft varying from relatively large airliners to the small F-104 fighter, under two conditions of binocular stabilization. Under one condition, the binoculars were fixed in a support that was aimed at a sector of the sky just above the horizon. Under this so-called "lay-on" technique, the higher power binoculars tended to increase detectability. In contrast, when hand-held higher powered and moderate powered binoculars (e.g., 6X24 and 8X22) were compared to unaided observation, it was found that the aided observations did not produce detection ranges greater than unaided observations.

As far as is known, no additional formal tests of optical aiding were conducted until 1971, when HumRRO resumed its study of the use of optical aids for target detection. Baldwin (7) employed a 1,000-to-1 reduced-scale simulation of an aircraft-detection situation. The objective of these laboratory studies was to evaluate the effectiveness of wide-field-of-view, low powered optical systems for initial target acquisition.

Observers were positioned 11 meters from a white background screen against which were presented black spherical targets that subtended less than 1 minute of solid angle to the observer. In this laboratory simulation, comparisons were made of target detectability under aided observation conditions using 7X50, 7X35, 2 1/2, or 3X optics. The luminance level of the screen at the position of the observer varied between 40 and 80 foot-lamberts, approximating the luminance levels of overcast to dark daylight conditions.

This research showed that under the low illumination conditions target detection was facilitated by the use of the 7X50 binoculars in comparison with the lower powered optics. Supplementary field tests that compared the 3X optic vs. unaided vision using similar target presentation conditions, but with approximately 1,500 foot-lamberts of sky background, again failed to show any significant facilitation because of using the 3X optics.

EFFECT OF EXHAUST FUMES

The only formal test of the effect of exhaust smoke upon aircraft detectability was conducted by the Department of Defense and the report in which it is discussed is classified. However, incidental observations have been made by other experimenters in this area. In the Gila Bend test reported by Wokoun (1), the greater detectability of the F-100 aircraft was attributed to the greater density of its exhaust fumes compared to those of the T-33. Similarly, Wright (3) concluded that the greater detectability of the F-4C aircraft in the head-on aspect could be attributed to the greater density of exhaust fumes emitted by the aircraft. A limited number of trials were also conducted in the German field test in which the F-104 was flown with and without the afterburner. The detectability of that aircraft was increased by approximately three kilometers when the afterburner was operating.

LOCATION OF THE OBSERVER

In only two of the field experiments was there an attempt to evaluate the effects of offsetting the observer from the flight path of the aircraft. In the Dona Ana test reported by Wright (3), the observers were located at three positions: no offset (i.e., the aircraft flew directly overhead at the termination of the flight), 650-meter nominal offset, and 1,400-meter nominal offset from the minimum crossing range. Nominal offset values equaled the average of the minimal offset distances of the aircraft that approached the observer groups at different headings from trial to trial.

Wright reported that offset did not affect the detection range of propeller aircraft, but did increase the detection range for jet-powered aircraft when optical aids were used. For unaided viewing, there was an inverse relationship between aircraft detection range and the offset of the observer. Wright attributed this result to the increased detectability of jet aircraft because of the smoke trails produced by the engine. As the offset increased between aircraft and observers, the presented area of the smoke trail increased. Wright conjectured that observers initially detected the smoke trail of the aircraft, thereby reducing the sky area to be searched. As a result, the aircraft itself was detected at a greater range by observers having larger offsets when viewing aircraft with binoculars.

The offset of the observer was also evaluated in the Tonopah tests reported by Frederickson *et al.* (4). One of the tests used the F-4C aircraft flying at low altitudes from a far distant horizon, with the observers located at OPs 200, 1,400, 2,600, and 3,300 meters from its flight path. This test indicated that aircraft detection range increased as offset increased, up to approximately 2,600 meters. Aircraft detection range varied between 9,800 meters for the 200-meter OP and 11,600 meters for the 2,600- and 3,300-meter OPs. Although the terrain elevation with respect to the ground projection of the aircraft's flight path varied between 5 meter for the 200-meter OP and +12 meters for the 3,300-meter OP, this variation in the elevation of the observers was minimal with respect to the variation in the aircraft detection ranges obtained. In contrast, it is highly likely that the observers having greater lateral offsets were initially detecting the smoke trail of the F-4C in a manner similar to that inferred by Wright.

SEARCH PATTERNS AND TRAINING

Most of the research conducted on visual search since the end of World War II has been concerned with the evaluation of patterns and techniques for air-to-ground or air-to-sea search. Very little research has been conducted concerning patterns of searching

sky from a ground observer's position. At first thought, it would seem that research on air-to-sea or air-to-ground search methods would find application in the ground-to-air search situation. Review of the research studies, however, indicates that the major concern of the investigators and analysts of the air-to-surface search problem has been with developing or designing optimal techniques for flying aircraft over large terrain surfaces to maximize the likelihood of target detection.

Techniques of searching sky from a ground position, however, have recently been studied by Baldwin (8). Using a simulation of a ground-to-air search situation, Baldwin evaluated two techniques of structured search compared to unstructured or untrained observation. These experiments compared several variations of a vertical zigzag or saw-tooth technique of searching a large display, as well as a horizontal technique of scanning. It was found that structuring the search method assisted some observers in detecting simulated aircraft, but paradoxically the same structuring technique interfered with target detection for other observers.

Informal observations made during this testing suggested that observers with relatively high visual acuity also tended to develop efficient search techniques through experience. There were indications that attempts to modify the search techniques used by these individuals produced a degradation in the effectiveness of the scanning operations. These observations suggest that persons with relatively poor visual acuity also do not possess efficient search or scanning techniques. Perhaps such persons would benefit from training in the use of systematic search methods, particularly those which tend to extend, and or capitalize on, scanning and search techniques developed through training and experience in reading printed material (i.e., a horizontal zigzag method of scanning the horizon). The research published in this area, however, suggests that techniques which partition the total area to be scanned into three or more subsections and which use systematic scanning of each subsection do increase detection likelihood, or reduce the acquisition time of hard-to-see targets.

Chapter 2

AIRCRAFT RECOGNITION ABILITY

Two characteristics of the recognition and identification responses are critical in a military situation: the range of the target at the time the recognition judgment occurs, and the accuracy of the judgment. In this chapter, various aspects of these response characteristics are discussed. In addition, since the military situation often involves a number of people working together, factors that affect individual and crew accuracy will also be examined.

RECOGNITION RANGE

BACKGROUND

The initiation of defensive action against an aircraft primarily depends upon a decision that the aircraft presented to the observer has a hostile identification. Unless it is unequivocally known that all aircraft in a theater of operation are friendly or that all are hostile, the identification response has traditionally been assumed to be dependent upon mediating recognition judgments. In the case of the man-ascendant air defense weapons, such recognition judgments are accomplished by weapon system operators, or crewmen.

Prior to the past decade, the artillery type of weapon systems available for defense against low-flying aircraft had relatively short effective ranges—1,500 meter or less. During the past 10 years, however, advances in guided missile technology have made it possible to develop defensive weapons that have much more extensive effective ranges against low-flying aircraft.

REQUIREMENTS

Because of these extensions in the effective range of defensive weapons, considerable interest developed within the military to determine how soon or rapidly an operator or crewman can make aircraft recognition judgments. That is, there was a desire to match the technical capabilities of the defensive weapon with the perceptual capabilities of the operators. In addition, there was considerable interest in finding ways to extend the recognition ranges of operators beyond that characteristic of unaided observation of aircraft. As a result, requirements arose for information on the relationship between recognition decisions and the distance between the aircraft and the observer.

GILA BEND TESTS

The earliest recognition range experiment was conducted during the Gila Bend test reported by Wokoun (1). Wokoun's recognition experiment was conducted in conjunction with the aircraft detection study reported in Chapter 1. Wokoun's observers all had visual acuity of 20/22 or better and were trained by the use of silhouette cards to recognize the

T-33, F-86, and F-100 aircraft. The visibility during his field experiment exceeded 15 miles and all aircraft were presented against clear sky. The recognition judgments were made following the detection judgments for each observer. Aircraft altitudes of 500 to 1,500 feet were employed with an aircraft speed of approximately 400 knots.

The results of the Gila Bend tests were presented as cumulative probability curves for all combinations of four search sectors (360, 180, 90, and 45°) and the two altitudes (500 and 1,500 feet). When summed over all these conditions, 50% of the recognition judgments, both correct and incorrect, occurred by the time the aircraft was 1,000 yards from the observer. Approximately 10% of the recognition decisions occurred when the aircraft was 3,000 yards or farther from the observer.

Wokoun's data indicated that, to some extent, recognition range was dependent upon the combination of search sector involved and altitude of the aircraft. For search sectors greater than 90°, the width of the sector and the altitude of the aircraft did not have great influences on the range at which the aircraft was recognized. However, for search sectors of 90° or less, aircraft tended to be recognized sooner (i.e., at greater distances) when at lower than at higher altitudes. That recognition range tended to be related to search sector may be partly a result of the relationship between detection range and search sector. Because recognition judgments follow detection responses, any factors that delay detection will necessarily delay recognition judgments.

DONA ANA TEST

The distances at which various aircraft are recognized was also included in the Dona Ana test reported by Wright (3). This field experiment concerning recognition range was also accomplished in conjunction with the detection field studies reported in Chapter 1. Of principal interest are the recognition range data obtained for the three jet aircraft (F-4C, F-100, and T-33), which flew at speeds of approximately 400 knots and at altitudes between 100 and 300 feet. One-third of the observers made recognition judgments without optical aids, and two-thirds used 6X30 binoculars.

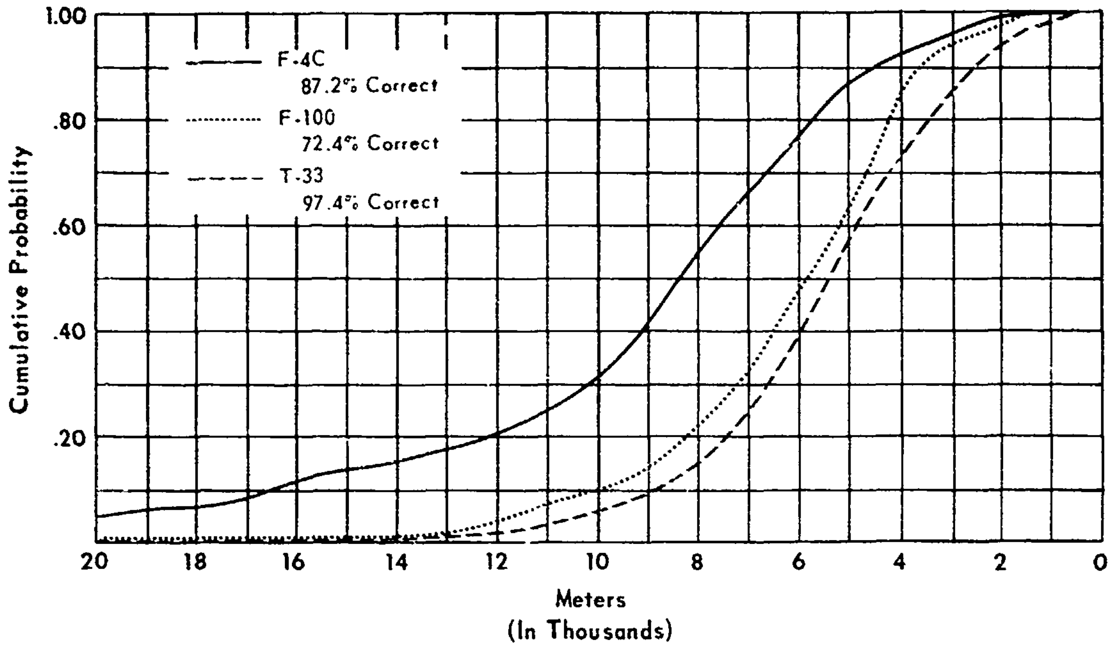
The 27 observers used in this field experiment were given both classroom and field training in aircraft recognition. The classroom training consisted of eight hours of recognition instruction on the aircraft to be used in the tests. This training utilized 35mm slides of models of the jet aircraft, using both the standard military techniques for giving aircraft identification instruction and the experimental method under evaluation.

In the field, the observers viewed 27 jet passes before actual testing began. For most of the practice trials, the type of aircraft to appear was announced prior to its initial pass. On other trials, only feedback on the type of aircraft was provided. After completion of the classroom training, observers correctly recognized 73.8% of the jet aircraft shown on a 40-item test. No proficiency test was administered to the observers following the additional field training.

In the Dona Ana Test, the observer was instructed to make (a) a tentative recognition response when he believed he could make a decision that was subjectively better than chance and (b) a positive recognition decision when he was subjectively "certain" he was correct. Although Wright obtained recognition range information for helicopters and propeller aircraft as well as for the jet class, only the jet data will be presented here. The cumulative frequency distributions of aircraft recognition judgments as a function of aircraft distance for the tentative recognition response for each of the three jet aircraft are presented in Figure 5. Similar data for the positive recognition judgments are shown in Figure 6.

The frequency distribution curve for each aircraft contains approximately 300 responses. The tentative recognition judgments were accurate 86% of the time, and the

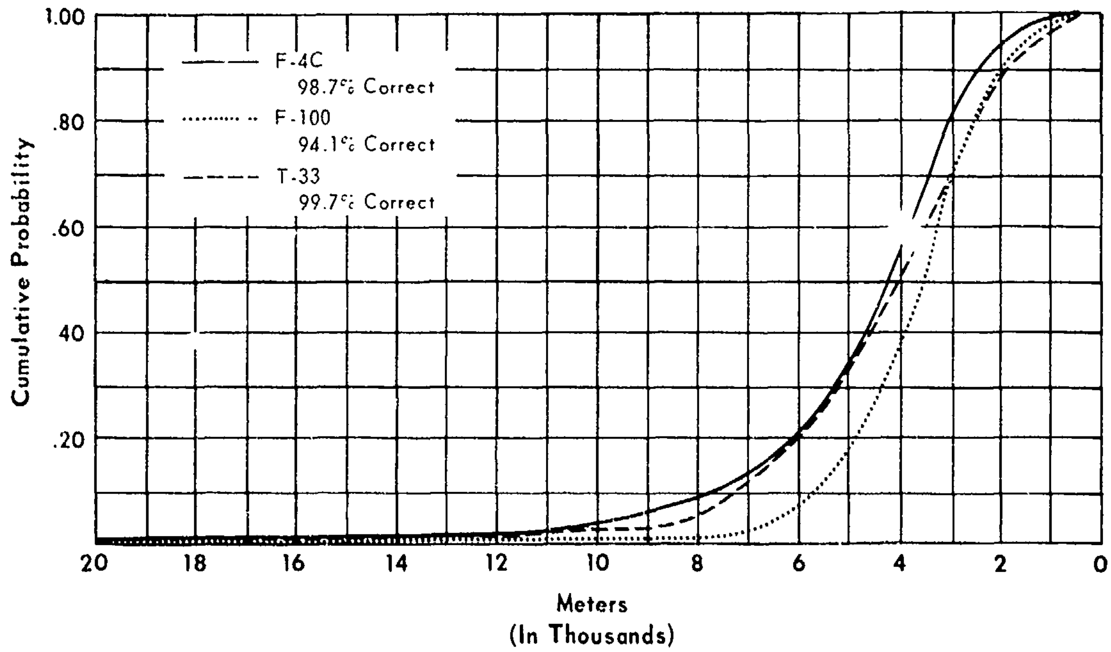
Empirical Probabilities of Tentative Recognition for Three Jet Aircraft



NOTE: From Wright (3).

Figure 5

Empirical Probabilities of Positive Recognition for Three Jet Aircraft



NOTE: From Wright (3).

Figure 6

positive judgments 97.5% of the time. When averaged over all observer offsets and viewing conditions (aided vs. unaided observation), the jet aircraft were recognized 50% of the time at distances greater than 4,000 meters. The recognition decisions occurred at approximately 5,500 meters or greater 25% of the time.

The distance at which the aircraft were recognized, however, depended upon viewing conditions and the observer's offset. Both the tentative and positive recognition ranges increased as offset increased, and, on the average, the jet aircraft were recognized at greater distances with binoculars than without them. However, it was noted that under the conditions used in Wright's test, binoculars tended to reduce detection distance of the potentially most threatening targets (i.e., targets having a zero offset or head-on presentation). Similar results also tended to occur for the unaided observations. That is, although head-on targets tended to be *detected* farther away (except when binoculars were used), the *recognition* of such targets was delayed beyond the distances characteristic of the more obliquely oriented flight paths.

Wright's data indicated that recognition judgments occurred at much greater distances than reported by Wokoun. The increased recognition distance may be attributed to several factors, including (a) Wright's observers had a smaller scanning sector than Wokoun's, and (b) Wright's observers apparently were more proficient in aircraft recognition than Wokoun's observers.

The two studies reported by Wokoun and Wright constitute the only known full-scale aircraft recognition field experiments that have been publicly described. Other tests, however, have been reported that used reduced-scale simulation of aircraft recognition situations in which model airplanes were used.

MINIATURIZED TESTS

REDUCED-SCALE SIMULATION STUDIES

The use of miniaturization, or reduced-scale simulation, of aircraft recognition situations was necessitated by the obvious difficulty of obtaining a wide variety of tactical aircraft for full-scale field studies. As has been noted, the full-scale field research programs were unable to obtain or use more than three jet aircraft. Because of the limited variety of aircraft used, the results of those studies have been criticized as not providing a basis for valid inferences to a typical air attack situation.

In 1967, HumRRO conducted tests of aircraft recognition using 1/72-scale models of aircraft in a miniaturized recognition situation (9). One pilot test used models of aircraft that had been used in the full-scale field study reported by Wright (3). The models were immobilized on a stationary pole and the observers were moved toward them from an initial distance at which recognition was impossible. When the recognition data were compared with the full-scale results from 1965, it was found that the statistical relationships between recognition frequency and target distance were fairly comparable for the two tests, after adjusting for the miniaturized scale factor that was used. However, the results indicated that for the miniaturized test the recognition judgments tended to occur somewhat earlier (target farther away) than for the full-scale test. This was attributed to the use of statically positioned models, which permitted longer periods of observation at each viewing distance.

Later in 1967, a more extensive miniaturized field experiment (Baldwin *et al.*, 9) was conducted that also used 1/72-scale model aircraft representing six U.S. and U.S.S.R. tactical fighters. Relevant parameters were scaled to 1/72-level, including the movement of the targets toward observers. Although the primary purpose of the field experiment was to evaluate the use of various communication sequences for aircraft recognition, the

experiment also provided an opportunity to test the validity of miniaturization, or reduced-scale simulation, of aircraft recognition situations. The results of the reduced-scale experiment, compared with similar data for the previous full-scale field study, are presented in Figure 7. A statistical evaluation of the two cumulative frequency distributions indicated that the two curves were not significantly different. It was concluded, therefore, that the reduced-scale test did provide valid estimates of the results that would have been obtained if a full-scale experiment had been conducted. Although this field experiment suggested that reduced-scale miniaturization provided a feasible method of evaluating aircraft recognition performance, no additional experimentation employing this method was accomplished for several years.

Results of Full-Scale and Miniaturized Recognition Range Tests

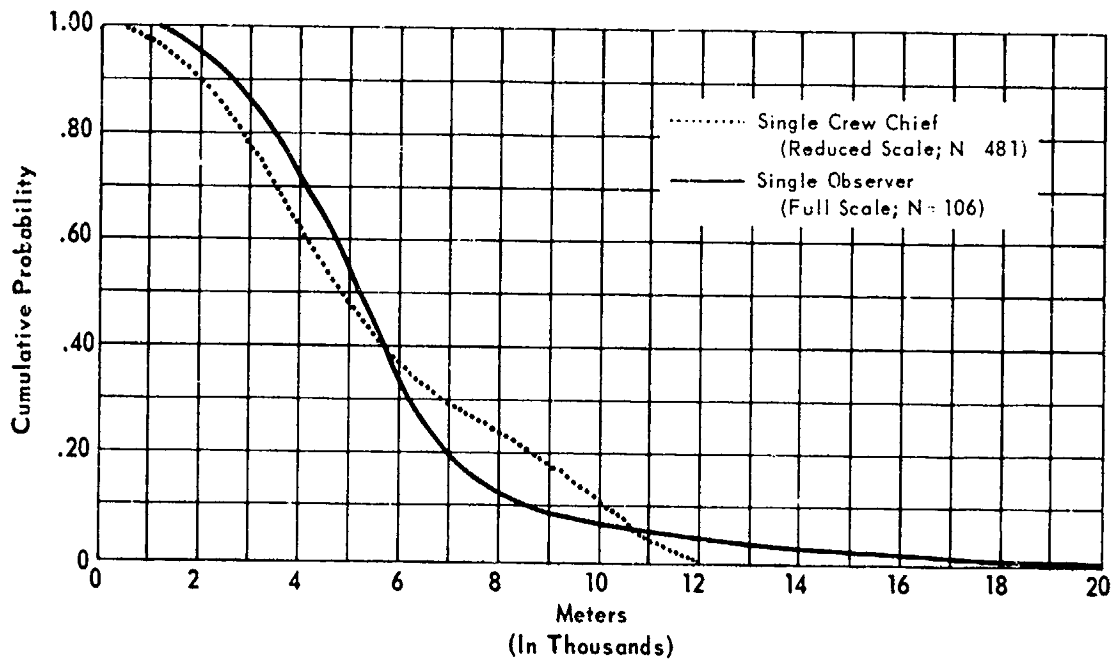


Figure 7

In 1971, HumRRO conducted another miniaturized field experiment, in an outdoor test facility near El Paso, to study the relationship between aircraft recognition range and the size, aspect angle, and color of the target (Baldwin, 7). This experiment employed a more extensive variety of targets in that eight single engine and five multiengine jet attack aircraft were simulated through the use of models. A 1/72-scale reduction of aircraft size, altitude, and approaching speed was used. Six aspect angles, or views, were used, ranging from a head-on presentation, through various oblique views, to a side, or 90°, orientation of the target. All observations were made by members of the military and the research staff, who were equipped with seven-power binoculars. These observers made approximately 3,000 recognition judgments over a nine-day experimentation period. Only 1.2% of these judgments consisted of uncorrected recognition decisions (i.e., errors). This high degree of accuracy attested to the correspondingly high level of recognition training attained by these "professional" observers.

Observers were located at one end of a scaled 16,000-meter flight path. The model aircraft were individually attached to a short boom mounted to the roof edge of a panel truck. The transporting vehicle moved at a scaled speed of 400 knots from the starting position. Each observer was provided with a hand-held reaction-time button and a response-choice box that he used when making his recognition decisions. Recognition judgments could be corrected at any time during a trial. Averaged over all aircraft of a similar classification (i.e., single vs. multiengine), recognition distance ranged from a minimum of approximately 7,800 meters (full scale) for the head-on view to a maximum of approximately 14,000 meters for multiengine aircraft with a climb of 45° and a heading of 35°.

It was noted that the average recognition ranges obtained in this miniaturized field experiment were much greater than those reported for previous full-scale and miniaturized tests. In fact, these average recognition ranges were similar to aircraft *detection* ranges obtained in previous studies. The increase in recognition ranges was attributed primarily to the higher skill levels of the professional observers used for this miniaturized study. All previous field tests, both full-scale and miniaturized, had used military personnel who were assigned to the experiment. The increased recognition range in this study was also attributed to increased target discriminability because of the object-to-background contrast level use.

COLOR (REFLECTANCE) EFFECTS

In previous field and reduced-scale experiments, aluminum-colored aircraft had been used. In this miniaturized field experiment, the models were painted a dark gray, which yielded high contrast with the sky background.

Additional testing was conducted in the reduced-scale situation to evaluate the effect of contrast ratio (reflectance) on recognition range. The supplementary test used two sets of six model aircraft, one set painted silver and the other dark gray. The silver models had approximately the same reflectance as a full-sized aluminum-skinned aircraft. The dark gray models had a reflectance about the same as aircraft painted with terrain camouflage colors. When averaged over all six aircraft, the mean recognition range for the dark-gray models was approximately 11,000 meters, compared to about 9,000 meters for the silver models. The overall difference between the two colors was statistically significant.

AIRCRAFT SIZE

As part of the miniaturized field study, reported by Baldwin (7), the size (area in square feet) of each of the aircraft at each of six aspect angles was determined by projecting silhouettes of each view against graph paper and measuring the area occupied by the silhouette. The correlation between presented area and mean recognition range for each of the 13 aircraft was determined for each aspect angle. Five of the six correlation coefficients were statistically reliable and ranged between 0.5 and 0.7. For some unexplained reason, the correlation between area and recognition range was not of sufficient magnitude for statistical reliability for the aircraft view associated with a 15° climb and 45° heading. These results suggested that 25% to 50% of the variability in recognition range can be predicted from a knowledge of aircraft size for most of the aircraft views. However, it was apparent that the presented area of the aircraft was not the major determiner of the time at which the recognition judgment occurs.

TRIAL-TO-TRIAL CONSISTENCY

The absence of a significant size-range relationship for the view with a 15° climb and 45° heading prompted additional analysis concerning observer reliability or consistency. In this context, reliability was concerned with an observer's trial-to-trial variation in the range at which each of the 13 aircraft was recognized. This consistency was examined separately for each aspect angle. Analyses and summarizations of the trial-to-trial variation indicated that, with the exception of the view with a 15° climb and 45° heading, each observer was quite consistent in the different ranges at which the 13 aircraft were recognized. However, the reliability of the recognition ranges for aircraft presented at the 15°-45° view was relatively inconsistent, particularly for two of the four observers.

CONSISTENCY ACROSS VIEWS

The stability of the recognition range for each aircraft across the different views was also examined. This analysis showed low consistency in recognition ranges across views. The most notable exception to this general finding was the set of correlation between views for head-on (0°-0°) vs. slightly obliquely oriented aircraft (10° climb-15° heading).

These analyses of sources of individual reliability suggest that observers employ different cues for different views of an aircraft, but for each specific view they are comparatively consistent in the relative distances at which various aircraft are recognized.

INTEROBSERVER CONSISTENCY

Analyses were also made of the consistency in the judgments made by different observers. Observers were paired and the average recognition range for each aircraft was correlated between pairs of observers for individual aircraft views. Among the set of six aircraft views, the least consistency occurred for the recognition ranges obtained by different observers for the head-on aspect. In contrast, the most consistent recognition ranges occurred for the view of the aircraft with the 45° climb and 35° heading. These results tended to support the analyses of the average recognition ranges, which indicated that the easiest view to discriminate was 45°-35° and one of the most difficult views was 0°-0° (i.e., head-on).

INDIVIDUALS VS. CREWS

BACKGROUND

The standard operating procedures for engaging low-altitude aircraft requires the crew chief to accomplish the visual recognition and identification function. For some air defense weapons, certain crewmen are assigned the task of forward observers (FO). These FOs are stationed somewhat remotely from the weapon, so that they can provide early warning of an impending attack. The FOs can also communicate their judgment concerning the aircraft identification to the weapon crew chief.

The addition of FO teams to air defense weapon crews could offer both potential advantages and disadvantages for system effectiveness. The more obvious potential advantages include the following:

(1) The FO could assist in the initial visual detection and acquisition of penetrating aircraft.

(2) The FO communication could provide a source of probable recognition before the aircraft was visible to the crew chief.

(3) The earlier FO recognition judgments could reduce the crew chief's uncertainty concerning identification and thus reduce system reaction time.

The potential disadvantages include the following:

(1) The time required for the communication between FO and the crew chief could increase system reaction time.

(2) The prior recognition by the FO, if incorrect, might predispose the crew chief to incorrect decision.

Previous field studies had evaluated the effectiveness of the recognition performance of observers working alone. There was no information concerning the recognition accuracy and reaction time of lone observers vs. observer teams. In 1967, HumRRO conducted an experiment to evaluate these two approaches to the recognition task (Baldwin *et al.*, 9). This experiment was conducted in an outdoor environment, using a 1/72-scale simulation of a recognition situation.

Prior to the test phase, 60 enlisted men were given classroom aircraft recognition training for three U.S. aircraft and three U.S.S.R. aircraft. The men were trained in groups of 20, one group on each of three successive days. The duration of training varied between four and eight hours, depending upon the rate at which the trainees achieved a 95% accuracy level for six views of the six aircraft. A reduced-scale field test was given two to four days after training. On each trial, a 1/72-scale model of one of the aircraft was mounted on a boom and moved at a simulated speed of 400 knots and at a simulated altitude of 200 feet.

Forty of the trainees were assigned to 10 four-men crews, consisting of a crew chief, a forward observer, and two communications assistants. The communications assistants were assigned to accomplish the communication operations for the FO and the crew chief. For one-third of the trials, the target passed over both the forward observer and crew chief positions, which were separated by approximately 460 meters (full-scale). For another one-third of the trials, the aircraft passed between the locations of the FO and crew chief positions, and for the remaining trials the aircraft passed at a scale distance of 2,500 meters offset to both positions. For one-half the trials, the crew chiefs received tentative and positive recognition judgments from the FO teams, and for the remainder of the trials the crew chiefs operated alone.

This study employed two performance measures, (a) recognition accuracy, and (b) remaining engagement time (RET), defined as the amount of time that elapsed between a crew chief's positive recognition judgment and the arrival of the aircraft at the minimum distance from the crew chief's location.

RECOGNITION ACCURACY

The results of the test indicate that the crew chiefs working alone were slightly, but not significantly, more accurate than when they received the FO judgments. However, these overall results obscured a number of crew chief-FO interdependencies that apparently occurred. Although the average accuracy of the individual chief was greater than the average accuracy of the FOs, the accuracy of the chief when he was working with a crew tended to be affected by the FO's accuracy. The research results suggest a number of specific instances in which the FO apparently either aided or hindered the crew chief's decisions.

REMAINING ENGAGEMENT TIME

When averaged over all trials, the remaining engagement time (RET) for crew chiefs working with FOs was slightly, but not significantly, larger than when these chiefs worked alone. However, the analysis showed that the 10 chiefs did not perform consistently under the two manning conditions—six chiefs made their decisions sooner when working with crews than alone, and four behaved in the opposite manner. Variation in RET also occurred in association with the offset of the target from the observer's position. When the flight path was parallel to the observer line and offset by 2,500 meters (scaled), all crew chiefs tended to make their judgments at comparable times. However, the two groups of chiefs mentioned above did not make decisions at comparable times for either the overhead or the intersecting flight paths—six of the 10 chiefs made their judgments eight to nine seconds before the others.

The results also suggested that there was a trade-off between recognition accuracy and RET when the chiefs operated alone. That is, when working alone, the more accurate crew chiefs tended to delay their judgments, while the less accurate chiefs made earlier decisions. Paradoxically, the FOs did not behave in this manner. The more accurate FOs tended to make early judgments, while the less accurate FOs delayed their decisions. Apparently the results of these opposing patterns was a tendency for the chief's accuracy to be positively related to the remaining engagement time when he was working with a crew. However, the chief's accuracy tended to be reduced when he worked with a relatively inaccurate FO. Six of the crew chiefs had an average accuracy of 88% when working alone and 89% when working in a crew. The remaining four chiefs, however, had an accuracy of 87% when working alone but 78% when working in a crew.

Additional analyses were made of the communication sequences that occurred between the FOs and crew chiefs. The test results suggest that the chiefs who did not perform as well when working with a crew as when working alone were more cautious in their decision-making than the more "effective" crew chiefs. The more effective crew chiefs frequently tended to make their final recognition judgments without waiting for the FO inputs. The less effective crew chiefs tended to behave in the opposite manner and, consequently, could have been adversely influenced by their FOs' inaccurate and/or late decisions.

Chapter 3

AIRCRAFT RECOGNITION TRAINING METHODS

This chapter contains a review of World War II studies of training procedures, particularly the WEFT and Renshaw methods of teaching recognition. Recent applied and analytical research that has produced substantial changes in training concepts for aircraft recognition also will be discussed. Discussions on (a) classroom techniques and (b) self-study and small-group methods of teaching aircraft recognition will be presented.

WORLD WAR II METHODS

BACKGROUND

Vicory, in a review of aircraft recognition training methods (10), reports that systematic studies of aircraft recognition training were first pursued in England in 1940 when that nation was being threatened by air invasion from Germany. According to an earlier researcher in this area (J.J. Gibson), the psychological theory existing at that time could not provide any clear guidelines with respect to effective techniques for teaching object recognition. As a result, the methods initially employed were based upon expert opinion rather than on systematic analyses of the nature of the recognition procedures and the development of such skills. As reported by Gibson, the British approach to recognition instruction included provision of information concerning the nature and characteristics of different military aircraft as well as the simple visual cues associated with shape and size.

The initial approach taken by the British could be considered object-analytic, in that considerable emphasis was placed upon analysis of the component parts of aircraft with the resulting development of a complex terminology needed to describe the shape characteristics of these components. As a result, the method that evolved for teaching aircraft recognition became known as the "WEFT" technique for describing the wings, engines, fuselage, and tail components. In this training technique, aspects of the shape of aircraft components that could be given names (such as "swept-back wings," "negative dihedral," and "leading-edge taper") were considerably overemphasized. There was also heavy emphasis on verbal learning; it was desired that lists of discriminating characteristics would be associated with the various visual images.

WEFT VS. TACHISTOSCOPES

In 1941, the WEFT system was adopted for training in aircraft recognition by the U.S. Navy and the U.S. Army Air Corps. In 1942, Samuel Renshaw of The Ohio State University proposed a radically different approach to teaching aircraft recognition. Renshaw's technique primarily emphasized the recognition of training images presented at extremely short (tachistoscopic) time intervals. The use of tachistoscopic presentation was based upon a hypothesis that brief presentations forced the observer to respond to the

total form of the image rather than to the aggregate of its component parts. According to Harvey's review of aircraft recognition training procedures (11), the advantages claimed for the Renshaw system of training were as follows:

(1) Rapid flashes forced instantaneous recognition.
(2) The use of rapid exposures of images gave the trainee experience in the sort of perception that he would need under combat conditions.
(3) Rapid flashes forced the trainee to see the total form of the plane, rather than the pieces and parts.
(4) Recognition of slides presented at short exposure durations was actually easier for the trainee than analysis of the images into its components.
(5) The use of rapid flash speeds increased motivation and the degree of attention to the material presented.

Basically, the method Renshaw proposed involved presenting images of the aircraft in brief flashes on the screen until the observer was able to identify it accurately. The durations were usually about 1/25 of a second. Subsequently, the exposures were gradually reduced to 1/75 or even 1/100 of a second on the assumption that these increasingly shorter intervals during training would yield a higher proficiency level. Harvey reviewed a number of previously unavailable World War II research studies concerning aircraft recognition, particularly as they pertained to the Renshaw method of instruction (11). He described studies conducted by R.M. Gagne and J.J. Gibson that refuted all the advantages claimed for the Renshaw technique. Although these studies were reported in 1944, the Navy and Army Air Corps had adopted the Renshaw method as early as 1942 or 1943, and this technique was used for the duration of World War II.

One of the logical bases for the Renshaw system was attacked by Gibson. Renshaw had contended that the use of rapid exposure of slides gives the trainee experience in the kind of perception he would need under combat conditions. Gibson pointed out, however, that in the usual combat situation the observer recognizes the aircraft long before it reaches effective firing range. He argued that, since the observer had considerable time to make recognition judgments, accuracy was more important than speed.

In this context, Gibson conducted a study in which observers were trained under three tachistoscopic intervals. One-third of the trainees were given instruction at an exposure duration of one second per slide. A second group was presented images lasting 1/10 second, and a third group had intervals lasting 1/50 second. All trainees were tested after instruction, using motion picture and slide presentation proficiency tests. The results showed no differences in the proficiency of recognition among the three groups trained with different exposure durations. During the 35mm test, the slides were presented at one second, 1/10 second, and 1/50 second for all groups. The slides shown for one second were more accurately recognized than those shown for the shorter intervals. In fact, the highest scores obtained by all three groups on the test slides were those associated with the images presented for one second.

A second experiment reported by Gibson showed that emphasizing aircraft features during the early days of training produced better aircraft identification performance than when the features were not emphasized, particularly for those distinguishing components that permitted discrimination between similar aircraft.

Harvey (11) also described an experiment reported by Gagne that directly compared the Renshaw and WEFT techniques. Two groups of 90 men each were taught the same aircraft in the same order using slides exposed for 1/10 second. In one group, instruction was given only on the total form of each aircraft (i.e., the Renshaw system) with no mention of features - such as shapes of wings, engine, or tail. In the second group, a standard set of distinctive features was emphasized for each aircraft presented. After 30 hours of instruction, both groups were tested with 45 slides of the 40 aircraft that had been included in the training program. The results of the proficiency test slightly favored

the WEFT system of presenting recognition instruction and did not support Renshaw's claims for the superiority of his total-form approach to training aircraft recognition.

In retrospect, it is unfortunate that much of the earlier research on recognition training has not been available to the public until recently. Its earlier availability would perhaps have lessened the debates that still occur among proponents of tachistoscopic techniques for teaching aircraft recognition.

CURRENT APPROACHES TO RECOGNITION TRAINING

THE SARGEANT METHOD

In 1956, the British introduced another technique for teaching recognition as an alternative to the WEFT and Renshaw methods. It was known as the Sargeant system for its originator, Charles Sargeant, editor of the *Joint Services Recognition Journal*. The Sargeant technique borrowed from both the WEFT and Renshaw approaches to recognition training. Although it did not use tachistoscopic exposures, the Sargeant method did use whole-image learning, and also emphasized learning the distinguishing features of aircraft. According to a British psychologist who evaluated the technique, it was believed that the distinguishing features were learned only in relation to the whole aircraft (Allan, 12).

The training materials employed in the Sargeant method for teaching a group of aircraft consist of two booklets for each aircraft to be learned. Initially, the aircraft are grouped by expert judgment according to the similarity of their design. The first book, which provides cues for identification, contains named photographs of different views of each aircraft and three plan-view silhouettes; this is the "key" material for making comparisons with the aircraft shown in a second book. The second book contains 120 to 140 target views of the same aircraft. After studying the aircraft features in the "key" book, the trainees attempt to recognize each aircraft in the second book. This process is continued until each aircraft is recognized correctly. Each trainee works alone, at his own pace and without formal instruction.

Allan conducted an experiment to compare the effectiveness of the Sargeant and WEFT systems. This experiment indicated that the Sargeant system was superior to the WEFT method in producing aircraft recognition accuracy.

A variation of the Sargeant technique is used currently in the *Joint Services Recognition Journal*. On one page of the journal, a few relatively large images of an aircraft, or models thereof, along with its silhouettes, are shown, while the opposite page shows a montage of many images of the same aircraft, along with a few "ringers." The student's task is to identify the "ringers," based upon their dissimilarity with the key images, as well as the other images in the montage. It is not known whether any systematic evaluation of this approach for teaching recognition has been made.

HumRRO's GOAR METHOD

In 1965, HumRRO initiated a research program to improve upon the WEFT-Renshaw methods of training aircraft recognition. Although the psychology of perceptual learning had advanced somewhat since World War II, there were still no firm psychological principles upon which to base a scientifically grounded technique for training form recognition. As a result, the initial approach developed by HumRRO, Ground Observer Aircraft Recognition (GOAR), although based on laboratory findings, was eclectic in its underlying principles and pragmatic in its objectives. In other words, all efforts were

made from the outset to produce an effective technique for training aircraft recognition, irrespective of theory or the lack of theory.

The GOAR method emphasized discrimination learning of the aircraft features that were relevant to identification requirements in a tactical situation and at tactically realistic distances. To facilitate learning to discriminate among similar aircraft during training, aircraft of similar characteristics were displayed simultaneously as well as successively (i.e., one at a time). The GOAR approach consisted of the following activities:

(1) Goal setting, which consisted of measuring trainees prelearning proficiency.
(2) Aircraft familiarization, which included familiarization with the nomenclature of features and provision of printed silhouettes of the three-planned views used as supplementary training aides.
(3) Discrimination learning by paired-comparison training involving the simultaneous presentation of pairs of images of different aircraft.
(4) Single-image recognition practice.
(5) Proficiency testing.
(6) Remedial instruction as needed.

In experimental form, the GOAR training program covered 16 aircraft that were grouped into four sets based upon expert judgments of their similarities. The training program consisted of 16 periods of 50 minutes each, which included paired-comparison training, successive image practice and review, proficiency testing, and remedial training if needed. More complete details on this training procedure are presented in a report by Whitmore, Cox, and Friel (13).

Initially, the GOAR method was compared with the WEFT-Renshaw technique. Two groups of trainees were administered one or the other type of training. After training, each group was given a proficiency test that consisted of aircraft views not included in the training program. The aircraft views in the proficiency test were also substantially smaller than those used for training. The test results showed that the average identification accuracy for those trained by the HumRRO technique was 61%, compared to only 20% for those trained by the WEFT-Renshaw method. Although this experimental program did not achieve the proficiency levels desired for an operational training technique, the method did show superiority over the procedures (WEFT-Renshaw) prescribed at that time for accomplishing aircraft recognition instruction.

The developmental effort described by Whitmore *et al* (13) should not be considered a formal experiment, but rather a comparison of two different approaches to instruction in aircraft recognition. The training program that used the WEFT-Renshaw training techniques was based upon existing U.S. Army instructional guidance on the administration of recognition training. That training approach was used by the research staff in order to ascertain the level of proficiency that could be expected from the existing instructional doctrine. In contrast, the modified program developed by HumRRO was administered to determine whether aircraft observers could be trained to a 95% level of accuracy, and also to determine the amount of time required to attain that goal. The experimental program developed by HumRRO did achieve a 95% accuracy goal for the aircraft views presented during training, but the average training time per aircraft was approximately 130% greater than that for the personnel receiving the WEFT-Renshaw approach.

The recognition training method described by Whitmore *et al* consisted of a state-of-the-art attempt to achieve a set of learning objectives through the use of laboratory-based pedagogical principles. Although the program produced the desired achievement level, the contribution of each of its components toward the attainment of the desired goal had not been separately evaluated in an aircraft recognition context prior

to the development of the training "package." Several studies subsequently were conducted in an effort to evaluate some of the more critical components of the training method.

TEACHING ONLY FRIENDLY OR HOSTILE AIRCRAFT

The Whitmore *et al.* (14) report also described a study designed to evaluate the effectiveness of limiting instruction in aircraft recognition to either friendly or hostile aircraft. Two experiments were conducted. In the first, approximately equal numbers of enlisted men were given recognition training on either six U.S. or six non-U.S. aircraft; but neither group was shown any other aircraft during training. Both groups were tested on all 12 aircraft upon completion of instruction. After all trainees had completed their instruction and had satisfied a 90% accuracy criterion, they were assembled as a group and administered the end-of-training test. The test consisted of seven views of the six aircraft on which they had been trained plus seven additional views not used in the training program. It also included 14 views of six aircraft that had not been included in the training program. Thus, the criterion test consisted of 168 images, evenly divided between familiar and unfamiliar aircraft, and further subdivided into familiar and unfamiliar views.

The amount of training time required to reach the 90% achievement level was compared for the two training conditions. Those given instruction on non-U.S. aircraft averaged 2.5 training sessions; those trained on U.S. aircraft averaged 2.0 sessions. The difference between the training durations approached statistical significance.

The recognition accuracy on the end-of-training test was also analyzed. Accuracy for familiar aircraft was approximately equal for the two groups, although there was a significant interaction between the training condition used and the class of aircraft presented during the criterion test. The students who were trained only on non-U.S. aircraft correctly identified 77% of the unfamiliar aircraft as friendly (U.S.), while the students trained only on U.S. aircraft correctly classified 66% of unfamiliar aircraft as hostile (non-U.S.). In other words, the students trained only on friendly aircraft incorrectly classified 34% of the U.S.S.R. aircraft as friendly.

A second experiment included paired-comparison discrimination between the two classes of aircraft. Under one condition, the students were told only that the non-U.S. aircraft were to be considered as hostile. For the second class, the trainees were also told the name (type designation) of each of the non-U.S. aircraft. A third group of trainees, which served as a control, received paired-comparison training involving only U.S. aircraft. They did not observe any non-U.S. aircraft during the training program. All three groups were instructed to learn the type of discriminations of the U.S. aircraft. The instruction was similar to that used in the first "friend-foe" experiment. Following instruction, all students were administered a proficiency test that included the U.S. and non-U.S. aircraft presented in familiar and unfamiliar views and that was scored for identification accuracy only. That is, any confusions in type naming within a nationality class were not scored as errors.

The results of the test indicated that the three training conditions were comparable as far as identification accuracy of friendly aircraft was concerned, and not significantly different in the identification accuracy of hostile aircraft. Average identification accuracy of friendly aircraft was approximately 88%, and of non-U.S. aircraft 49%. The control group from the previous day was given additional training on a second day, which consisted of paired-comparison and successive-image instruction involving non-U.S. aircraft only. This group was readministered the proficiency test and the performance of these

students on the first and second administrations was compared to evaluate the effectiveness of the additional instruction on non-U.S. aircraft. For U.S. aircraft, the seven students were approximately equally accurate on both days. For the non-U.S. aircraft, the average accuracy increased from approximately 50% on the first day, which followed instruction only on U.S. aircraft, to 82% on the second day after instruction on the non-U.S. aircraft. This increment was statistically reliable.

The results of these experiments indicated that: (a) when only one group of aircraft is included in training, the accuracy of identifying unfamiliar aircraft was unacceptably low, and (b) in order to achieve desired proficiency levels following instruction in identifying aircrafts of all classifications, such aircraft must be included in the training program.

ANALYTIC STUDIES OF CLASSROOM METHODS

A series of studies reported by Whitmore *et al.* (14) had the following objectives:

(1) To identify the minimum number and type of aircraft views used during training that would produce a uniformly high level of recognition transfer to all aircraft views of operational significance.

(2) To establish an operationally valid time interval for exposing aircraft images during training.

MINIMUM TRAINING VIEWS

In a series of transfer-of-training studies, trainees were instructed with limited sets of aircraft views and then tested on a wider variety of recognition views. Seven different configurations of training views were studied, varying from a single view of an aircraft to a maximum of nine views. In each of these experiments, the trainees were required to learn to recognize six aircraft. The proficiency level desired upon completion of instruction varied between 80 and 90% from study to study.

Two experiments used a single view of each aircraft for instruction, and two others used three views (one training condition used the three plan-form training views traditionally used in recognition training). A fifth experiment presented five views of each aircraft during instructional periods, and the final experiment presented nine views of each aircraft. The views for the proficiency test consisted of training views as well as images not used in the training program. A total of 30 different views was shown for each of six aircraft during the end-of-training tests. The effectiveness of each training condition was evaluated with respect to the recognition accuracy for familiar vs. unfamiliar views.

The average recognition accuracy for familiar views varied between 79 and 91%. The accuracy for views not used in training varied between 54 and 81%. Minimum decrements in recognition accuracy occurred for the training conditions that used the greatest number of views for instruction. A minimum loss of 5% resulted from the use of the nine training views. The maximum decrements were associated with the training programs that used a small number of views (either one or three). Moreover, students trained with the three plan-form views were no more accurate than students trained with only a single oblique view. In contrast, the training condition that used three obliquely oriented training views produced a greater level of transfer of training to unfamiliar views than did the training with the plan-form views.

This series of experiments indicated that the views used during training must be selected to provide uniform generalization across all aircraft views of tactical significance.

EXPOSURE DURATION

Whitmore *et al.* (14) also studied the effect upon recognition accuracy of varying the exposure interval during testing. A group of students was trained by a common instructional method, and then divided into three subgroups. Each subgroup was tested with different image exposure durations—one, three, or five seconds. The results of this experiment indicated that the duration of exposing the images during testing did not have a significant influence upon recognition accuracy. The generality of this finding, however, is limited to the conditions of this experiment—specifically, to trainees who had been taught to discriminate among six aircraft with an end-of-training accuracy of 90% or better.

SUCCESSIVE VS. SIMULTANEOUS PRESENTATION

In 1965, Gavurin (15) reported an evaluation of two methods of presenting aircraft during recognition training. The aircraft to be learned were presented successively, in the first condition and simultaneously during the second condition. Gavurin found that, on a subsequent proficiency test, significantly greater identification accuracy was achieved by people who had been trained with the simultaneous procedures.

PAIRING AIRCRAFT FOR DISCRIMINATION LEARNING

Presumably, a portion of the effectiveness of the Sargeant and GOAR techniques may be attributed to their provision for simultaneous comparison of images of different aircraft. Somewhat conflicting results have characterized research on this particular problem. Harvey (11) cites a study reported in 1944 in which it was found that recognition proficiency did *not* depend upon whether similar or dissimilar aircraft were presented together during training. Recently, however, Vicory (16) reported an extensive series of experiments in which the opposite results were obtained.

A series of experiments was conducted to evaluate various strategies for pairing different aircraft and different views during instruction (16). In one experiment, the discrimination practice consisted of simultaneous presentations of pairs of (a) highly similar aircraft or (b) aircraft with low similarity. On the end-of-training test, it was found that recognition accuracy was greater for those trainees who had been instructed with highly similar pairs of aircraft.

Vicory also evaluated four strategies for teaching discriminating cues or attributes by which aircraft can be distinguished. One technique involved teaching only the cues that discriminated between friendly and unfriendly aircraft. The second technique involved teaching discrimination between aircraft within a nationality class, the third included both types of discrimination training, while in the fourth, trainees received instruction in which no attribute training was given. The results indicated that either technique for teaching the distinguishing attributes was effective, although techniques that involved teaching *both* within and between class attributes tended to interfere with recognition accuracy on the proficiency test. These effects were more pronounced for discriminations between highly similar aircraft.

Two strategies for pairing aircraft images during discrimination training were evaluated in another of Vicory's experiments. In this experiment, the discrimination training consisted of presentation of either (a) the same view of different aircraft or (b) different views of different aircraft. After training, all subjects were given a proficiency test, the results of which indicated that identification accuracy was higher when the same views of

different aircraft were paired for discrimination learning during training than when different views of different aircraft were paired.

Another aspect of Vicory's experiments included an evaluation of the provision of rule pretraining (prior to formal discrimination learning). Rule pretraining consisted of instructing trainees in how to use rules to classify and combine aircraft attributes for making recognition decisions. Vicory's experiments involved learning to recognize four aircraft, two U.S. and two non-U.S. Thus, he was able to establish a four-classification response situation that depended upon the joint values of two independent binary dimensions. One binary, the affirmational rule, involved between-class attributes (i.e., attributes that distinguished the U.S. from the non-U.S. aircraft). Within each of these classifications a second binary decision, a conjunctive rule, was available to further distinguish between the name discriminations of each pair of aircraft.

The results of these experiments showed that rule pretraining had some effect for discriminating between similar views if no similarity cueing instruction had been given. However, when dissimilar views of different aircraft were provided for discrimination learning, neither rule pretraining nor the attribute training had substantial effects upon test performance—although performance did improve when the two were combined. The greatest inaccuracy on the proficiency test was produced by pairing dissimilar views of different aircraft in combination with no preliminary rule pretraining. Under this condition, a significantly greater number of between-class errors in the recognition judgments occurred. Rule pretraining substantially reduced between-class errors, but increased within-class errors.

SELF-STUDY AND SMALL GROUP TECHNIQUES

Aircraft recognition training has traditionally consisted of group instruction using projected slide images supplemented by individual study of silhouette cards or recognition sheets. Such training is normally given during formal individual and unit training periods. Analyses of military training requirements indicate, however, that such group approaches need to be supplemented by training materials suitable for self-study (or for very small groups) in a highly flexible training schedule.

In 1971, Miller and Vicory (17) reported a series of experiments designed to evaluate alternative programs for teaching aircraft recognition using printed visual imagery, rather than projected imagery. The primary objective of the research was to identify effective printed self-instructional training programs that could be used to supplement formal classroom instruction. Six different types of training materials were evaluated:

(1) Multi-Image Cards (MIC). Each of these cards pictured five different views of one aircraft, along with a brief description of its most distinctive features. The trainee's first task was to study each card to get a general concept of each aircraft. Then he compared similar views of different aircraft across the cards. The MIC was designed to be used early in training when similar aircraft should be compared for discrimination learning.

(2) Paired Comparison (PC) Cards. Each of these cards pictured two, or occasionally three, aircraft at the same view. Aircraft names were printed under the pictures on one side of the card, while the reverse side presented the same pictures without names. The student first studied the paired images along with the aircraft names, then turned over the deck of cards to practice naming the aircraft.

(3) Flash Card Drill. Each of these cards had a picture of an aircraft on one side, and the same picture plus the name of the aircraft on the reverse side. There was one card for each aircraft and view. After attempting to name an aircraft by inspecting

the card, the student turned over the card to reveal the correct answer. In using these materials, students were instructed to follow a "drop-out procedure" in which a card was eliminated from the deck after the aircraft view was correctly recognized. The flash card drill was designed for a more advanced discrimination learning level, since it required performance under circumstances much like the ultimate test conditions.

(4) Sargeant. The Sargeant procedure for training aircraft recognition has been described earlier in this report. With this procedure, the student first studied the "key book," which showed a few views of each aircraft along with a written description of its distinctive features. Next, the student attempted to identify the aircraft in a second (problem) book by referring back to the key book as needed. The Problem Book (Book II) contained the 60 views on which the student would subsequently be tested. This book had 120 items, each aircraft view being presented twice at various image sizes.

(5) Sorting. Each trainee was given a stack of 60 cards (one card for each view of each aircraft) and a sorting board that had six spaces, one for each aircraft. Above each space was a written description of the distinguishing features of an unnamed aircraft. The student sorted the cards into six stacks, one for each aircraft, using the verbal cues presented. On the second sorting, the name of each aircraft was also exposed and the student repeated the sorting procedure. On the second sort, the student could see both the cues and the aircraft names. A third sorting was conducted in which only the names of the aircraft were visible, but no cues were available for his assistance.

(6) Sorting Game. This was a competitive card game based upon the sorting procedure just described. The first "hand" was conducted using the sorting method, then the game element was introduced On the second hand, each man paired with another as an opponent and both played on a common board. The opponents had their cards in the same ordinal order, and turned over each card at the same time. The object of the game was to place each card in the correct space before the opponent. Bonus points were given for catching an opponent's mistakes.

The relative effectiveness of these various teaching materials was evaluated on the basis of an end-of-training test consisting of a slide recognition test previously used in HumRRO research. Progress was measured periodically during training by a printed version of a recognition test. Various combinations of the training materials were administered to seven groups of students.

A total of 135 men participated in the complete experimental comparison. The number of trainees assigned to each training condition varied between seven and 18. Because none of the trainees had previously received formal instruction in aircraft recognition, their motivation was not expected to be as high as that of men required to learn aircraft recognition for their military job specialties. The training period for each man was about half a day, and covered six aircraft.

The combination of procedures that were evaluated represented various feasible training programs, but no one group used more than four of the six types of training materials. For example, one group of 28 students began instruction with the Multi-Image Cards; moved to Paired Comparison and Flash Cards; practiced with the Sargeant materials; and were then administered the end-of-training test. A second group commenced instruction with the sorting task and then received practice with Paired Comparison and Flash Cards. For a third group, Flash Cards were not included in the sequence of instruction.

The least structured training program consisted of a sampling of all the materials. These students were instructed not to spend much time on any one procedure and time was called in each phase even though several men had not yet finished. This program was intended to give the students a sample of all the procedures and then administer the end-of-training test.

Based upon performance on the end-of-training test, the highest average proficiency level and the least amount of inter-trainee variation in accuracy was attained by Group 1, which received structured practice involving the Multi-Image, Paired Comparison, and Flash Cards, as well as a review using the Sargeant materials. The lowest level of achievement was attained by the group that started out with the sorting task and then practiced with Paired Comparison cards before taking the final test. When comparisons were made among the seven training conditions, it was found that Group 1 was significantly more accurate than each of the other groups, except the group with the training condition that heavily emphasized Flash Cards followed by sorting. In all instances, Group 1 was, however, characterized by a higher achievement level than any other group.

Additional analyses showed that the trainees in Group 1 who received Paired Comparison practice before Flash Card practice achieved higher levels of proficiency than the trainees who were administered these materials in reverse order (i.e., Flash Card practice following Paired Comparison training apparently was more effective than Paired Comparison training following Flash Card practice). Other analyses suggested that the final procedure used with Group 1 (the Sargeant method) probably could be eliminated, since there was no additional improvement on the periodic tests given during instruction after practice with these materials.

It was also found that Group 1 attained their learning goal in the least amount of time. This group averaged 96% correct recognition following an average training time of 71 minutes—or about 12 minutes per aircraft.

Chapter 4

RANGE ESTIMATION

INTRODUCTION

This chapter is concerned with several different aspects of distance estimation or range determination. Man's ability to estimate the distance to ground targets has been widely explored, but very little is known about the ability of an observer to estimate the distance to moving aerial objects.

Research on ground-to-air range estimation ability has been characterized by variations in the definition of the range or distance estimation task. In some research, the observers have been required to estimate the intervening distance between their position and a moving target at some random point in time (Wright, 3, Frederickson *et al.*, 4). Other research has been concerned with estimating the arrival of an aerial object at some specified criterion distance or range (McCluskey, Wright, and Frederickson, 18).

Still other studies have been concerned with determining stadimetric accuracy: that is, how accurately an observer can judge the match, or coincidence of, the apparent size of a moving target and some standard or reference object (McCluskey, 19). The latter task does not constitute a range estimation task per se. Stadimetric ranging requires the observer to judge the equality in either the vertical or horizontal substance of two objects, one of which is changing in its substance (in this case, a moving aircraft). Stadimetric ranging involves matching apparent sizes rather than estimating the distance that intervenes between an observer and some distant object. Since range estimation or distance determination can be accomplished by stadimetric ranging techniques, research on the latter method of estimating an object's distance is included in this discussion.

DEFINITION OF RANGE ESTIMATION ERROR

In common parlance, people may talk about overestimating a distance of underestimating a size. In research on distance estimation abilities, there have been conflicts in the ways in which the terms "overestimation" and "underestimation" have been used. In research conducted by Wright (3), for example, if an aircraft at a distance of 15,000 meters was judged by an observer to be at 12,000 meters it was said that the observer *underestimated* the distance to the target. On the other hand, if the aircraft was at 10,000 meters and the observer said it was at 15,000 meters, it could also be said that he *overestimated* the distance between his position and the target.

In contrast, in studies reported by McCluskey *et al.* (18), some experiments required the observers to estimate when an aircraft had reached some specified (or criterion) range, such as 1,500 meters. For studies involving the estimation of specific predesignated criterion ranges, errors in estimation are calculated as the actual distance to the target minus the judged (i.e., criterion) distance. For example, if an observer is required to estimate an open-fire distance of 800 meters, and he does so when the aircraft is actually at 1,000 meters, McCluskey *et al.* would say that the observer had *overestimated* the amount of distance consumed by 800 meters; if the observer signaled that an aircraft was

at the criterion distance when, in fact, it was at a shorter distance from him, they would say he *underestimated* the amount of distance consumed by 800 meters.

For the sake of consistency, the definitions used by McCluskey *et al.* for "overestimation" and "underestimation" will be used in this report. This usage has been selected because most of the research on ranging has been concerned with observers' accuracy in judging when specific criterion distances or events have occurred. Relatively little research has been devoted to evaluating the accuracy of individuals in judging a wide variety of intervening distances between themselves and moving objects.

The research on stadimetric ranging can also be confusing with respect to the use of over- and underestimation. When an observer is attempting to match the sizes of a dynamically changing object and a fixed or stationary object, he can, for example, overestimate either (a) the size of the dynamically changing object or (b) the size of the static object.

ESTIMATION OF AIRCRAFT DISTANCE

This portion of the report is concerned with the ability of observers to judge the distance intervening between their position and the location of an object. For example, an observer might be asked to judge the distance between himself and a barn or to the top of a tree. As applied to the aircraft ranging situation, he would be asked to estimate the distance between himself and the position of an aircraft at one or more times during its flight path. For example, we might ask, "As soon as you detect it, tell us how far way it is." Or, "As soon as you can identify the aircraft, please estimate its range." Wright (3) included distance estimation in the extensive Dona Ana field test described in the detection and recognition sections of this report.

ACCURACY WITHOUT FEEDBACK

Wright's observers were given preliminary training in range estimation, which consisted mainly of practice in estimating distances varying between 350 and 2,000 meters to ground targets, using the size of familiar objects as ranging aids (e.g., fence-posts). As described by Wright, the major purpose of this training was to provide the observer with a basis for establishing or developing a "reasonably calibrated yardstick" to use in making his estimates during the field tests. In the field test itself, however, there was no opportunity to provide feedback or target range information to the observers.

During the field test, each observer made three distance estimation judgments during each flight of a target, subsequent to the detection, tentative recognition, and positive recognition responses. Two-thirds of the distance estimation judgments were made without use of visual aids, and one-third while the observers were using binoculars for the detection and recognition judgments.

Observers were located at three positions with respect to the aircraft's flight path: directly underneath, 650 meters offset, and 1,400 meters offset. The results of this field test suggested that both the use of binoculars and the observer's offset did affect the magnitude of the ranging errors. In Wright's study, ranging errors were computed as estimated range minus actual range. Therefore, a distance was *overestimated* if the observer said the aircraft was farther away from him than it actually was. The ranging errors varied between large overestimates for those observers located directly under the flight path to large underestimates for observers located on offsets from the flight path. The most accurate distance estimations were made by the group with the 650-meter offset.

The nature of the errors also tended to vary as a function of the actual distance to the aircraft. For distances beyond 3,000 meters, the aircraft's location tended to be overestimated, whereas underestimates characterized range estimates for the target positions of 3,000 meters and less, *if* the observers were located under the flight path or slightly offset. However, for the observers located at the 1,400-meter offset, all target distances tended to be markedly underestimated (i.e., aircraft erroneously judged to be nearer). Ranging errors also tended to be smaller for observers who used binoculars.

ACCURACY WITH FEEDBACK

Distance estimation was also included in the field studies reported by Frederickson *et al.* (4). In this field test, observers were positioned along a line perpendicular to the flight path at observation posts (OPs) 200, 1,400, 2,600, and 3,300 meters from the crossover point. Jet fighter aircraft flying at very low altitudes (less than 200 feet) were the targets. The observers estimated the slant range from their location to the aircraft at a signal given by test control personnel. The actual ranges to be estimated varied between 1,000 and 5,000 meters, with the maximum slant range increasing as the OP's offset from the flight path increased. Both minimum and maximum ranges varied according to the observers' location from the flight path. For the 200-meter OP, their range varied between 1,000 and 4,000 meters. For the most distant observer location, the 3,300-meter OP, the true slant ranges varied between 3,400 and 50,000 meters.

This task involved distance estimation, rather than estimation of the criterion range, because the observers did not know when the test control center would order a distance estimation judgment to be made. However, the observers did know the minimum and maximum limits of the slant ranges that would be characteristic for their OP. In this test, it was possible to provide observers with knowledge of results concerning the accuracy of their judgments. Immediately after the aircraft reached the crossover point, the test control director informed the observers of the actual flight-line distance to the aircraft from the crossover point at the time the judgment was required. The observers were provided with a conversion table for each OP, which permitted them to transform flight-line distance to the corresponding slant range for the OP. Thus, they were able to determine the magnitude of their error in judging the distance to the aircraft, although this procedure did create a delay in the feedback provided them.

Differences in the type, or direction, of error tended to vary among the offset positions. At the 200-meter OP (which might be considered almost under the path of the aircraft), the observers consistently tended to underestimate the true slant range. For example, they estimated an aircraft that was actually 4,000 meters away to be, on the average, 3,200 meters distant. The results of this field test indicated that the average error decreased as the observers' offset increased.

EFFECT OF OBSERVER OFFSET

There was also a tendency for the variability among observer judgments to decrease with increasing offset (Frederickson *et al.*, 4). The average error for the 200-meter offset was −176 meters (an underestimate). The average algebraic errors for the other three observer locations were −81, 29, and +53 meters. For the almost head-on observations (from the 200-meter OP), the magnitude of the errors increased as true slant range increased. At the more laterally positioned OPs, the average errors included overestimations as well as underestimations; as a result, there was no trend for error magnitude to change systematically as a function of true slant range.

The results of this test were also analyzed in terms of an error dispersion index (DI), which was defined as the square root of the second moment about 0 error:

$$DI = \sqrt{{}_0\mu_2^2} \quad \text{or} \quad DI = \sqrt{M^2 + \sigma^2}$$

This index provided a measure of the total dispersion of the estimation errors about the true slant range and reflected biases (that is, consistent errors in judgment among observers) and error variation caused by individual differences in judgments. The DI tended to increase in magnitude as the range to be estimated increased. In addition, except for the observations made at the 200-meter offset, the average judgmental bias (average error) contributed little to the total magnitude of DI. These results indicated that the major portions of the estimation errors were due to variations in judgmental accuracy both within and between the various observers.

FULL-SCALE TRAINING METHODS

COMPARISON OF PAIRED-ASSOCIATE, FEEDBACK, AND STADIMETRIC METHODS

The two major field studies (3, 4) provided both a foundation and a stimulus for additional studies to evaluate factors that may influence distance judgment errors. Several experiments were reported by McCluskey *et al.* (18) that were concerned with accuracy in judging criterion ranges such as 400 or 1,500 meters. In one study, the observer's task was to estimate when an aircraft had reached a distance of 350 meters. Observers were to make two 350-meter estimates, one when the aircraft was inbound and one when it was outbound. The observers signaled the 350-meter event by depressing a pushbutton. This study of criterion range estimation evaluated two techniques of training men to make such judgments and a third technique involving the use of stadimetric methods of determining distance. The aircraft flew at a constant speed of 100 knots and altitudes of 175, 300, and 400 feet. The trainees were given 18 practice trials in estimating the 350-meter range for inbound and outbound aircraft.

One group was taught by a paired-associate training method. In this method, the instructor announced slant ranges of 450 meters diminishing to 250 meters for both incoming and outgoing directions of the aircraft. Students were told to attend to the apparent size of the aircraft and to try to remember its appearance at a criterion distance of 350 meters.

A second training method involved immediate reinforcement or feedback. The trainees in this group were informed that as the aircraft passed over they would be told to make a distance judgment at two different times—one as the aircraft was incoming and one as it was outgoing. On command, the trainees were to estimate the distance to the aircraft. Immediately after they had recorded their answers on a score sheet, they were informed of the correct distance of the aircraft at the time they made their estimate. A random sequence of aircraft distances, varying between 250 and 450 meters, was used during the training trials. Two-thirds of the judgments during practice were required when the aircraft was either at 300, 350, or 400 meters slant range.

A third group of observers employed stadimetric ranging techniques, using their index finger held at arm's length as a ranging aid. Using this finger occlusion technique, the third group observed the flight of the aircraft and heard announcements of the slant ranges as the aircraft moved toward and away from their position. Their task was to try to remember how much of the aircraft was occluded by the index finger when a range of 350 meters was announced.

Prior to receiving any training, each of these groups had been administered a preliminary test to evaluate their baseline performance level. After training, there was a statistically significant difference between the pretest and an end-of-training test for all three groups. The average error during the pretest was +229 meters for all conditions and both directions of flight, whereas the post-training test had an average error of +53 meters. There was, however, a significant difference between the error magnitude for the incoming and outgoing directions, which occurred for both the pretest and the post-training test.

The magnitude of this error varied with the training condition. The students who had received either the immediate reinforcement or the paired-associate training had an average incoming error of approximately 150 meters. The direction of this error was defined as an overestimation since the observers apparently erred in their judgment of the amount of space covered by 350 meters. The average outgoing error was approximately −50 meters, that is, an underestimation. In contrast, students trained on the stadimetric ranging method (finger occlusion) had average errors varying between +68 meters for inbound flights and −60 meters for outgoing flights. When averaged over both flight directions, this training technique and job procedure had the smallest net error of the three techniques evaluated.

ALTITUDE AND ILLUMINATION

A second experiment was later conducted to evaluate the effects of two factors that could, potentially, influence estimation accuracy—aircraft altitude and amount of illumination at the eye. The aircraft flew at approximately 100 knots at two altitudes, 75 feet and 400 feet, corresponding to target elevation angles of 9° and 55° above the horizon. Two levels of illumination at the eye were achieved by the use of variable density goggles. Performance under normal daylight conditions (approximately 1,500 to 2,000 foot-lamberts in the southwestern United States) was compared to a condition of reduced illumination that approximated an overcast day. This was achieved by adjusting the variable density goggles to near maximum polarization. The light transmitted through the goggles was then approximately 5% of the ambient illumination.

The observers were required to estimate when an aircraft had reached a distance of 350 meters from their position. The observers used in this test were the same as those who had received the three different training and/or aiding techniques described in the previous paragraph. The results of the experiment indicated a statistically reliable difference in the error magnitude for the two target altitudes. The mean error was +88 meters for low elevation and +32 meters for high elevation. The overall mean errors were +140 meters for incoming estimations and −21 meters for outgoing directions.

No significant differences were found between the two illumination levels. The researchers noted that the magnitude of the incoming-outgoing difference in error was less when the target was at high elevation, which suggested that aircraft elevation may have been partially responsible for bias observed in the previous tests.

USE OF HELICOPTORS IN TRAINING

Additional training experiments were conducted to obtain further evaluations of unaided estimates of criterion ranges. The earlier experiment had evaluated paired-associate and immediate reinforcement techniques using relatively slow aircraft and a criterion range of 350 meters during training and testing. The later research made further comparisons of these two training techniques for criterion ranges of 400, 800, 1,500, and

2,500 meters. This training experiment used jet aircraft at altitudes of approximately 100 feet.

Twenty-eight men participated, 14 for each of the two training methods. Each group was further subdivided into two sections. One section in each training method was given preliminary training involving the use of helicopters to provide a relatively stationary target for initial learning. Although a pretest was scheduled prior to the conduct of any training, instrumentation failure at the outset of the experiment resulted in a loss of the pretest data.

Following training, each group was given a post-training test on a terrain area different from that used during training. This was done because the results of an earlier full-scale field study (4) had indicated that terrain features may be used as cues by observers to determine criterion ranges, and it was desirable to eliminate familiar terrain cues (e.g., those that may have been used during training) for the end-of-training tests.

Each trainee was required to learn to estimate the four different criterion ranges. During the test, each trainee was required to make two estimates of one of the ranges each time the aircraft flew over—one for the incoming and one for the outgoing direction. The specific ranges required on a given flight were announced prior to the trial. The observers received extensive instruction, consisting of 36 training trials each day for three days, plus a daily proficiency test consisting of 12 additional trials.

This training experiment involved a combination of the characteristics of both criterion range and distance estimation tasks. Although the research objective was aimed at determining the effects of training methods on criterion range estimation, the training methods used required the observers to make distance estimations that varied about the criterion ranges. For example, the observers receiving the immediate reinforcement training were told to make an estimation, upon command, at two different points during the pass of the aircraft—one incoming and one outgoing.

The instructor gave a ready signal approximately two seconds before saying, "Estimate now." Upon hearing this command, the observers recorded their estimate of the aircraft's range at that instant. After the trial had been completed, they were told the correct range. One-third of the training trials involved estimates of the aircraft position when it was at one of the four criterion ranges. On two thirds of the trials, the "Estimate now" command was given when the aircraft was at greater or lesser ranges.

For the paired-associate training, a series of five consecutive ranges was announced for both incoming and outgoing directions of the aircraft. Each series consisted of one of the four criterion ranges accompanied by the bracketing ranges. Observers were requested to pay attention to the apparent size and distance of the aircraft as the ranges were announced and to "keep in mind" the ranges they were being trained to estimate (i.e., the criterion ranges).

For those men who received the supplementary helicopter instruction, one third of the trials on each day were conducted at a separate training facility, which used the helicopter as a target. The training procedures using the helicopter were the same as those used with the jet aircraft, except that the helicopter instruction involved observation of the essentially stationary aerial target. Students in the helicopter training groups were informed that the jet aircraft would be approximately three times as large as the helicopter; otherwise, the instructions for both the immediate reinforcement and paired-associate training methods were like those received by observers not having the supplementary instruction.

Analysis of the post-training test given on the final day showed no significant difference between the two groups trained with only the jet aircraft. A significant difference was found between the two training methods for the sections that had received the helicopter instruction. Those who had helicopter instruction in combination with

paired-associate learning tended to substantially underestimate the greater criterion ranges, whereas those who received the helicopter-supplemented, immediate-reinforcement training tended to overestimate these distances.

For shorter distances, there were no appreciable differences among training procedures, whether supplemented or not. In addition, there was an interaction between the criterion range involved and the flight direction of the aircraft. Although the criterion ranges of 400 and 2,500 meters tended to be underestimated for both flight directions, the other two criterion ranges (800 and 1,500 meters) were overestimated for incoming and underestimated for outgoing flights.

Aircraft altitude again had an influence upon error magnitude; the distances to aircraft flying at an altitude of 750 feet tended to be overestimated to a greater extent than for aircraft at 250 feet. Further evaluations of training methods suggested that the supplementary helicopter instruction may have tended to increase the accuracy of judging relatively short criterion ranges, but tended to produce estimation errors for 1,500 and 2,500 meter ranges. In general, the immediate reinforcement instructional method produced the smallest judgmental errors. This experiment suggested that approximately 100 training trials were required to learn to estimate all four criterion ranges with reasonable accuracy. However, since additional training was not provided, it was not known whether further increases in average accuracy and decreases in individual differences would have resulted if additional instruction had been provided.

TRAINING IN MINIATURIZED SITUATIONS

The use of miniaturized, or reduced-scale, facilities for training in distance and criterion range estimation was included in the studies reported by McCluskey *et al.* (18). A small-scale pilot study, using a 1:50-scale reduction of range estimation situations, was conducted in conjunction with the comparison of training methods reported in the previous paragraph. In the pilot study, five observers were trained by the paired-associate method to judge five distances that bracketed the four criterion ranges of 400, 800, 1,500, and 2,500 meters. A 1:50-scale model of an F-100 aircraft was employed as the ranging target. The model aircraft was fixed to the top of a stationary pole, and the observers walked toward or away from it as the various incoming and outgoing ranges were announced.

Following this training, which consisted of 36 practice trials, the observers were tested in a full-scale environment, and the results of their post-training tests were compared with the results from observers who had been trained in a comparable full-scale environment. Statistical analyses of the two groups (reduced-scale vs. full-scale training) indicated that the judgmental errors were consistent except for one distance, 1,500 meters inbound. For that test condition, the observers who had received reduced-scale training had systematically larger errors than those who had received full-scale training. With this exception, the pilot study suggested that reduced-scale training was a feasible method for developing skill in criterion range estimation. Even in this reduced-scale training situation, large estimation errors were associated with the movement of the inbound aircraft, similar to that which had occurred for observers receiving full-scale training by other methods of instruction.

The results of the pilot study stimulated a greater interest in the feasibility of employing reduced-scale, or miniaturized, approaches to training range estimation. A subsequent experiment, also reported by McCluskey *et al.* (18), used two methods of estimating criterion ranges, but used a reduced-scale training environment. In this experiment, the immediate reinforcement training method was employed, in conjunction with two approaches to reducing range estimation errors, particularly those associated with the

inbound or head-on view of aircraft. The pilot study indicated that the inbound estimates were greatly in error in that the observers tended to overestimate the criterion range required. One technique to overcome this constant error was to deliberately bias the training of inbound criterion range estimations, by employing a false or biased scale during the training program.

A second group of observers were trained to make the distance estimation observations while viewing the aircraft through a partially closed fist with about a dime-size aperture at the far end. This resulted in the observers using monocular vision in viewing the aircraft with a reduced field of view. The aperture also provided a stadimetric aid for comparing the apparent size of the aircraft with the diameter of the aperture.

Both groups of observers were also instructed to attend to the size of the aircraft and to pick out distinguishing features that were visible at the various training distances.[1] Both groups viewed the model monocularly during training. Again, the aircraft was stationary and the observers moved toward and away from the model. Each observer estimated each of the four criterion ranges, first inbound and then outbound. During each trial and following each estimation judgment, the observers were informed of the correctness of their response. The results of a subsequent full-scale test were analyzed statistically, and it was found that learning had occurred as a result of the training experiences for both groups. However, the average error for those using the stadimetric aid (closed fist) was significantly smaller than for the observers who were trained by the biased training method. Again, it was found that range estimation errors for the incoming, or head-on, aspects were significantly larger than for the outgoing direction.

USE OF THE RIFLE AS A STADIMETRIC AID

The final experiment in this series of training research studies involved determining how many training trials were required for observers to learn to estimate a single scale distance of 350 meters, while using a stadimetric ranging technique. The stadimetric aid in this case consisted of a replica of the front sight guard of a military rifle. All training was conducted on a miniaturized situation using 1/72-scale models of aircraft that were moved on an electrically powered carriage.

The model aircraft was moved at a scale speed representing 100 knots and the observers were stationed at distances scaled to represent 0-, 100-, 200-, and 300-meter offsets from the target path. The training method employed immediate knowledge of results, which followed each observer's judgment of the criterion range. In this situation, instrumentation was devised so that when the observer made a criterion range judgment, the aircraft movement was terminated and the actual distance of the aircraft from the observer was announced.

If the estimate was wrong, the target was then moved to the correct position to provide opportunities for additional observation by the trainees using the stadimeter. The accuracy of their judgments following training was tested in a transfer study, again using model aircraft, which included scale models of four tactical jets that varied in fuselage length and wingspan. The test results did not show any significant variation in error associated with the size of the aircraft. In this test situation, the outgoing range judgment errors were found to be significantly larger than the incoming errors. There was also a significant variation in accuracy as a function of the observer's offset.

The incoming judgments of criterion range were considerably more accurate than had been observed previously in full-scale and miniaturized studies. This increased

[1] That is, the *detection* of local cues of component parts is another possible technique for estimating distance.

accuracy was attributed to the use of an indoor testing situation, which could have provided visual "anchor points" for judging the relative distance from the observer to the target. Offset also influenced the accuracy of the judgments of criterion range. The judgments tended to be underestimates for the first three offsets (0, 100, and 200 meters) and then shifted to overestimates for the 300-meter offset.

Several additional criterion range estimation studies were conducted by HumRRO. McCluskey (19) reported on these studies of various aspects of stadimetric ranging techniques using reduced-scale facilities and immediate knowledge of results of error during the training program. The training was accomplished in a reduced-scale environment. End-of-training tests used full-scale test facilities with F-100 jet aircraft at altitudes of 150 feet and a speed of 400 knots. The training program used a 1/48 reduced-scale simulation of these parameters. During training, the model aircraft were mounted on horizontal bars affixed to and transported by a panel truck. The stadimetric ranging task was accomplished using shoulder-supported weapon mock-ups.

These studies were designed to evaluate training for one specific criterion range, 1,500 meters. Each weapon mock-up had a flat, black metal post at its terminal end that simulated the forward gun sight guard of a weapon. One experiment was conducted to evaluate whether the feedback provided during instruction was qualitative or quantitative information. Results of the full-scale testing indicated that the rate of learning was comparable for the two methods and the test performance showed no significant differences between the two training conditions.

A second experiment evaluated the accuracy of criterion range estimation in a full-scale environment following training in a reduced-scale situation. During training, observers were given qualitative information on the magnitude of their errors (i.e., they were informed whether they had fired too soon, too late, or correctly for 1,500-meter engagement distance). Following the instructional period, the trainees were given a reduced-scale criterion test, then tested, and retested 24 hours later, in a full-scale environment. The retesting was done to provide preliminary information on short-term retention of the accuracy of stadimetric ranging skills. Analysis of the test results indicated that accuracy varied over the tests and also as a function of the direction of flight. The average error for incoming targets was +10 meters across all tests, a slight overestimation of criterion range, whereas for outgoing targets, the error was −171 meters. Curiously, the magnitude of errors for both incoming and outgoing targets was essentially the same for the reduced-scale tests, whereas for the two full-scale tests, the errors were significantly less for incoming than for outgoing targets. On the average, the observers underestimated outgoing distances by approximately 200 meters or more on the full-scale tests.

RETENTION OF STADIMETRIC SKILL

A third experiment reported by McCluskey (19) was concerned with longer-term retention of stadimetric ranging training. The research design provided for reduced-scale training, followed by a reduced-scale test, and, 30 days later, by a full-scale field test. The full-scale field test was then supplemented by additional training and another full-scale test to evaluate the effectiveness of remedial instruction on error magnitude.

In the third experiment, 36 observers were trained in a reduced-scale situation. These observers were in training to be crewmen of a forward area air defense weapon in which ground-to-aircraft ranging was a requisite skill. Thirty-two of these observers were available 30 days later to take the full-scale retention test. For unexplained reasons, the refresher training used a paired-associate instructional method, whereas the original learning had involved the immediate feedback technique. The range estimation judgments

obtained in the reduced-scale test and the two field tests were analyzed to evaluate the effect of test conditions and the direction of movement of the target. When averaged over the three tests, the magnitude of the error was significantly smaller for incoming aircraft (+55 meters) than for the outgoing aircraft (-165 meters).

Although not discussed by McCluskey, the analysis also indicated a statistically significant interaction between the errors associated with flight direction and the repeated tests. Although both field skill tests produced larger judgmental errors than the reduced-scale test given immediately following training, the nature of the errors during the full-scale tests and during the reduced-scale test differed.

In the reduced-scale test, errors of judging incoming and outgoing criterion ranges were underestimations of the true criterion range. In contrast, the underestimation errors associated with the outgoing distance increased from approximately 75 yards, characteristic of the reduced-scale test, to over 200 yards for the first full-scale test, and to approximately 200 yards underestimation for the second full-scale test. Paradoxically, the judgmental errors for incoming aircraft shifted from a negligible underestimation for the reduced-scale test to overestimations of approximately 50 meters for the first full-scale test and more than 100 meters for the second full-scale test.

From these results, it appears that the refresher instruction given between the two full-scale tests tended to increase the judgmental errors for inbound targets. When the average estimation errors for the full-scale test administered 30 days following completion of training are compared with the estimation errors for the observers at the beginning of training, it is apparent that the judgmental errors associated with outgoing targets reverted to the levels characteristic of untrained observers. In each instance, the observers had average errors that underestimated the true criterion range by 150-200 meters.

Paradoxically, the judgments during the retention test of the criterion range of inbound targets did not revert to the error levels characteristic of the observers at the beginning of instruction. At the outset of training, the observers' average errors had fluctuated widely, from approximately 200 meters underestimation to essentially no error. The initial retention test, however, was characterized by a negligible average error in the estimation of criterion range for inbound targets.

GREATER ERROR FOR OUTBOUND FLIGHT

This tendency for estimation of outbound criterion ranges to be grossly in error and highly unstable was characteristic of several of the later experiments reported by McCluskey *et al.* (18) and McCluskey (19). The earlier experiments that involved variation of training methods and use of full-scale environment tended to produce overestimations of inbound targets and either accurate estimates or underestimations of the outgoing aircraft. The later studies that employed reduced-scale training and stadimetric aids for criterion range determination tended to be characterized by gross underestimations of the outgoing ranges, irrespective of the criterion range to be estimated.

Although stadimetric aiding tended to produce accurate estimates of inbound criterion ranges, these aids did not overcome the tendency for outgoing estimates to be greatly underestimated. For example, the last experiments reported by McCluskey for determining a criterion range of 1,500 meters were characterized by underestimations of outgoing range of 200-250 meters; that is, when the aircraft was actually 1,250-1,300 meters distant, the observer signaled it was at 1,500 meters.

LABORATORY SIMULATION OF STADIMETRIC TASKS

In 1972, Ton reported laboratory experimentation that attempted to determine the factors that cause underestimation of outgoing criterion distance.[2] These experiments used an electronic simulation of approaching and receding aircraft. The target was a narrow horizontal bar presented on a cathode ray tube. The bar targets were programmed to either increase or decrease in magnitude, and the rate of change in size (speed) could be varied.

Observers viewed the target monocularly through a one-inch aperture. An illuminated stadimetric aid having a horizontal subtense equal to the wingspan of an F-100 aircraft at 1,500 meters was positioned in a light-proof box between the observer's eye and the cathode ray tube display. Observers were instructed to signal the coincidence of the size (horizontal extent) of the variable target and the stadimetric standard.

Earlier, McCluskey (19) had hypothesized that the observed incoming-outgoing difference in accuracy would decrease with decreasing aircraft speed, and as the range to be estimated increased. This hypothesis was based upon an analysis of the rate of change of aircraft subtended angle as it approached or receded from the observer. For an inbound target and a criterion range of 1,500 meters, the rate of change of the visual angle subtended by the horizontal dimension of an aircraft changes relatively slowly and with a more or less linear rate prior to the criterion range. That is, the apparent acceleration of the aircraft is constant and low.

In contrast, for an outgoing target (receding stimulus size) the apparent size of the aircraft experiences more rapid decelerations prior to the criterion range of 1,500 meters. McCluskey speculated that the difference in rate of change of size of the target prior to the criterion distance influenced the accuracy of judging the coincidence event, since the observer presumably must anticipate the criterion (coincidence) event in order to minimize errors in signaling such coincidence. Since, for outbound targets, the rate of change in size of the aircraft is relatively great, it was hypothesized that the observers would anticipate the coincidence event earlier than they should or would if rate of change was linear or low. The relationship between change of target size and target distance is shown in Figure 8.

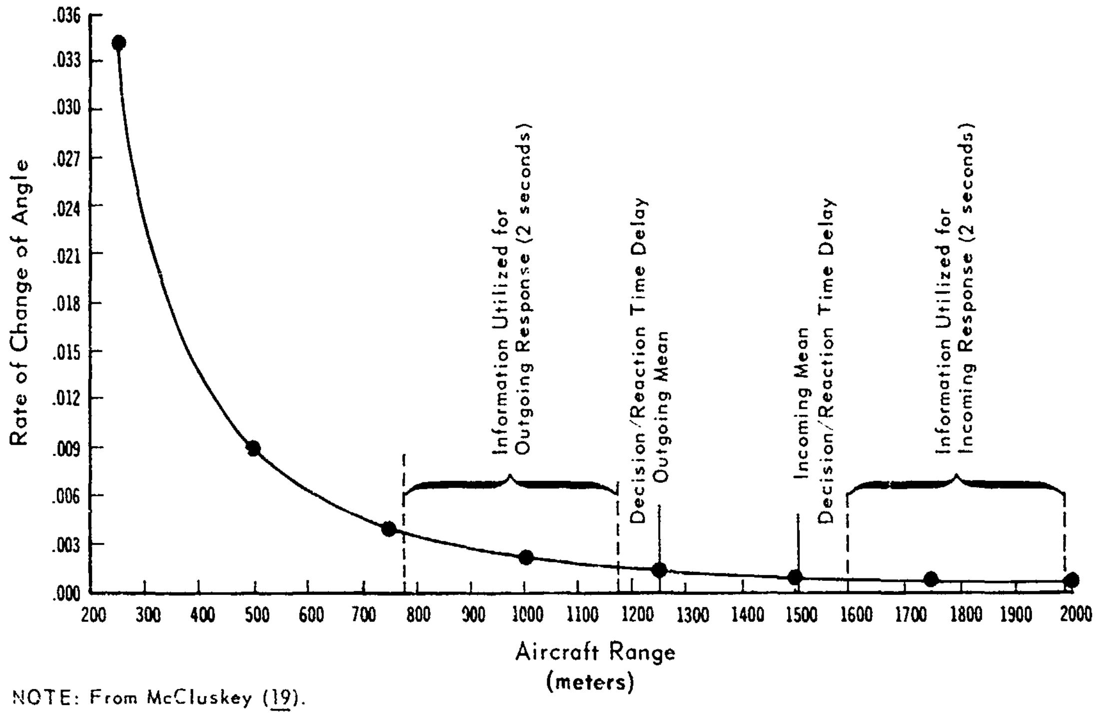
Ton conducted an experiment to test this hypothesis. He simulated target speeds at 200, 300, and 400 knots for approaching and receding aircraft. The direction of movement and speed were varied randomly over trials, and observers were instructed to signal the coincidence event for both approaching and receding target sizes.

Averaged over all target speeds, the accuracy of the coincidence judgments for inbound targets was quite high; the difference between the average error and the criterion distance was not significant. There was, however, a significant average error for the receding or outbound targets. On the average, judgments of the coincidence of the receding and standard stimuli erred by a time interval that would produce a 200-meter underestimation of criterion range. The magnitude of this error was comparable to that in previous full-scale field studies. Variation in target speed, however, did not influence error magnitude.

Since this approach to simulating the range estimation task did produce the differential error levels associated with ascending- and descending-sized targets, it may be inferred

[2] "An Investigation of the Sources of Error in Stadimetric Range Estimation," unpublished manuscript by William H. Ton, HumRRO Div. No. 5.

Rate of Change of Aircraft Subtended Angle as a Function of Range, Experiment II



NOTE: From McCluskey (19).

Figure 8

that the laboratory experiment did capture critical variables associated with full-scale criterion range or judgments. Assuming that the laboratory simulation provided a valid abstraction of a range estimation task, McCluskey's hypothesis concerning the relationship between error levels and aircraft speed is not supported by Ton.

Additional experiments are being conducted by the author of this report to evaluate the effect that target motion or movement has on the accuracy of such coincidence decisions. These studies will also examine the effect upon error rates of using different types of stadimetric aids, since it is possible that the results obtained by McCluskey and Ton are at least partially attributable to design characteristics of the device used for making the comparison decision. In at least a portion of the future experiments, dynamic target motion will be eliminated for some experimental comparison purposes, to be substituted with graduated changes in target size.

FACTORS THAT AFFECT ACCURACY

A review of the available literature on distance estimation, criterion range judgments, and stadimetric ranging (coincidence judgments) indicated that no research has found error-free ranging behaviors. Although training methods and other factors have been varied and relatively large numbers of subjects have been used (in some cases), the average estimations for a group of observers always deviate significantly from the true physical range or distance. Estimation errors associated with incoming vs. outbound, or ascending vs. descending-sized targets have been discussed. Earlier research by McCluskey *et al.* (18), which has already been discussed, included target elevation angle and ambient illumination level as potential factors that might influence ranging accuracy.

ILLUMINATION LEVEL

The one reported experiment on illumination level varied the amount of light reaching the eye by requiring observations to be made through polarized glasses. Even when the amount of light reaching the eye was attenuated by 95%, there was no effect upon ranging errors, as compared with normal daylight illumination levels (18).

TARGET ALTITUDE

The effect of target elevation above the horizontal plane has been examined for only a relatively short distance judgment—350 meters (18). After observers were trained to estimate a criterion range of 350 meters, using targets with high and low elevation angles and four different training procedures, the observers were tested for their accuracy in estimating this range for targets with elevation angles of 9° and 55° above the horizon. In this experiment, the average error magnitude associated with low targets (low visual angles above the horizon) was greater than for high altitude targets.

The results of this experiment also suggested that the difference in estimation errors associated with high and low elevation targets varied as the function of the method used for teaching criterion range estimation. The observers used in this experiment had been trained by several different training techniques: immediate reinforcement, paired-associate, stadimetric, and uncontrolled practice. Although this aspect of the results was not discussed by the authors (18), the estimation errors characteristic of the observers who were trained to use a stadimetric technique for criterion range estimation showed a pattern of errors that differed from that displayed by observers trained by the other three techniques. Except for those trained in stadimetric methods, the high elevation targets tended to produce larger errors in estimating inbound ranges than outbound, particularly under conditions of low ambient illumination.

Just the opposite results occurred for the finger occlusion technique. However, with this technique, the errors associated with low-altitude targets were greater for estimates of inbound than outbound targets. Since this experiment was conducted early in the research problem concerning the accuracy of distance estimation, the variation in the error pattern that occurred for stadimetric ranging apparently was attributed to experimental error or artifacts in the results. These early results, however, tend to be consistent with later results obtained by the same researchers concerning the relative accuracy of stadimetric estimates of outbound versus inbound criterion ranges.

ESTIMATION ACCURACY AND OTHER SKILLS

TIME ESTIMATION

In conjunction with Ton's research on the effect of target speed on the accuracy of criterion range estimation, he also conducted studies to identify visual skills that may be related to the accuracy of criterion range judgments. He hypothesized that accurate anticipation of the coincidence between variable and fixed stimuli was related to accuracy in an individual's ability to estimate time intervals. He hypothesized that correct anticipation of coincidence was based upon an accurate interpolation of target velocity based upon its movement history. Perception of target velocity, in turn, was hypothesized to be based on an internal integration of stimulus movement over time. That is, an individual who possessed an accurate "time-sense" would produce more accurate estimates of coincidence events of the type associated with stadimetric ranging.

Ton conducted two experiments that required the observer to duplicate the time intervals that elapsed between two auditory signals, using intervals of 8, 11, and 16 seconds. After the second signal was presented by the experimenter terminating the standard interval, the observer was asked to initiate an identical interval by pressing a switch and then performing a number cancellation task that was intended to prohibit internalized counting. When the observer judged that the time interval was equal to that presented by the experimenter, he again depressed a switch. The errors in duplicating the presented time intervals were subsequently correlated with range estimation errors. Neither experiment yielded a significant correlation between these perceptual measures.

FLICKER FUSION

Ton also investigated the relationship between range estimation errors and flicker fusion, another measure of perceptual time-sense. The flicker fusion frequency is the frequency at which an intermittent, or flickering, light is perceived as a steady light. Ton used a circular flickering field that subtended 2° of visual angle against an illuminating surround subtending 30° including a fixation point in the center of the flickering field. Illumination of both the flickering target and its surround was fixed at 40 foot-lamberts. With this apparatus, Ton measured the frequency at which the flickering light was perceived, by each observer, as a steady light. The converse transition was also determined and the average frequencies were computed. These flicker frequencies were then correlated with range estimation in two experiments, but neither experiment produced a reliable correlation between these two perceptual measures.

PERCEPTUAL STYLE

Ton also studied the relationship between estimation and other visual-perceptual measures, such as perceptual style and visual acuity, but again found no relationship between these measures and estimation errors.

Ton's results tend to suggest that variation among individuals in their ability to estimate criterion ranges is unrelated to other aspects of visual or perceptual efficiency. However, supplementary analyses of Ton's results, by the present author, leave open the possibility that variation in range estimation and accuracy *within* an individual is itself highly variable. That is, the results obtained by Ton for his criterion estimation task indicate that his various observers were relatively unreliable in judging the coincidence events.

A reliability coefficient of approximately .57 was obtained through supplementary statistical routines on Ton's analysis of variance results. Since the measure to be predicted, criterion range estimation, was not itself highly reliable, attempts to identify factors related to stable differences among individuals in this skill would be thwarted. Additional research concerning individual differences and ability to estimate criterion range seems to be needed, using observers who, through training and practice, have become stabilized in their ability to interpret their estimate of the coincidence event. Only after individual errors in coincidence event determination have become stabilized would it be possible to seek the identity of other factors that may be contributing to variation among individuals in their judgments of criterion range events.

LITERATURE CITED

LITERATURE CITED

1. Wokoun, William. *Detection of Random Low Altitude Jet Aircraft by Ground Observers*, U.S. Army Ordnance Human Engineering Laboratories Technical Memorandum 7-60, Aberdeen Proving Ground, Md., June 1960.

2. Zimmer, W.D. and McGinnis, C.F. *Redeye Target Detection Study*, Technical Memorandum 1072, Army Missile and Evaluation Directorate, White Sands Missile Range, N.M., March 1963.

3. Wright, A.D. *The Performance of Ground Observers in Detecting, Recognizing, and Estimating Range to Low-Altitude Aircraft*, HumRRO Technical Report 66-19, December 1966.

4. Frederickson, E.W., Follettie, Joseph F., and Baldwin, Robert D. *Aircraft Detection, Range Estimation, and Auditory Tracking Tests in a Desert Environment*, HumRRO Technical Report 67-3, March 1967.

5. Kurke, M.I. and McCain, Claude N., Jr. *Low Power Optical Systems and Aerial Target Detection*, Technical Memorandum 5-57, U.S. Army Ordnance Human Engineering Laboratory, Aberdeen Proving Ground, Md., June 1957.

6. Doetsch, K.H. and Hoffmann, H.E. *Die Maximale Auffpssreichwrite Bei Der Optischen Beobachtung Tief Aufleigender Fluszeuge Vom Boden Hus*, Braunschweig, Deutsche Forschungsanstalt fuer Luft-Und Raumfahrt, Institut Fuer Flugguehrung, 1966.

7. Baldwin, Robert D. *Relationship Between Aircraft Range and the Size, Color, and Aspect Angle of Aircraft*, HumRRO Technical Report in press, [1972a].

8. Baldwin, Robert D. *Attempts to Improve Visual Detection Through Use of Search Patterns and Optical Aids*, HumRRO Technical Report in press, [1972b].

9. Baldwin, Robert D., Frederickson, Edward W., and Hackerson, Edward C. *Aircraft Recognition Performance of Crew Chiefs With and Without Forward Observers*, HumRRO Technical Report 70-12, August 1970.

10. Vicory, Arthur C. *A Brief History of Aircraft Identification Training*, HumRRO Professional Paper 27-68, August 1968.

11. Harvey, L.O. *Survey of Visual Research Literature on Military Problems During World War II (Papers Collected by the Armed Forces - NRC Vision Committee)*, Research Paper P-453, Institute for Defense Analyses, Arlington, Va., April 1970.

12. Allan, Mary D. "The Sargeant System of Recognition as an Approach to the Learning of Perceptual Skills," *Occupational Psychology*, vol. 32, no. 4, 1958.

13. Whitmore, Paul G., Cox, John A., and Friel Don J. *A Classroom Method of Training Aircraft Recognition*, HumRRO Technical Report 68-1, January 1968.

14. Whitmore, Paul G., Rankin, William C., Baldwin, Robert D., and Garcia, Sandra. *Studies of Aircraft Recognition Training*, HumRRO Technical Report 72-5, February 1972.

Preceding page blank

15. Gavurin, E.I. *An Evaluation of Various Tachistoscopic and WEFT Techniques in Aircraft Recognition*, Naval Training Device Center Technical Report IH-40, U.S. Naval Training Devices Center, Port Washington, N.Y., 1965.

16. Vicory, Arthur C. "Empirical Comparisons of Selected Procedures for Aircraft Identification Training," Doctoral Dissertation, University of New Mexico, Albuquerque, N.M., August 1969.

17. Miller, Elmo E. and Vicory, Arthur C. *Comparison and Evaluation of Printed Programs for Aircraft Recognition*, HumRRO Technical Report 71-22, October 1971.

18. McCluskey, Michael R., Wright, A.D., and Frederickson, E.W. *Studies on Training Ground Observers to Estimate Range to Aerial Targets*, HumRRO Technical Report 68-5, May 1968.

19. McCluskey, Michael R. *Studies on Reduced-Scale Ranging Training With a Simple Range Finder*, HumRRO Technical Report 71-24, December 1971.